frontiers | Frontiers in Virtual Reality

TYPE Original Research
PUBLISHED 11 August 2025
DOI 10.3389/frvir.2025.1594350

OPEN ACCESS

EDITED BY
Justine Saint-Aubert,
Inria Rennes - Bretagne Atlantique Research Centre, France

REVIEWED BY
Adélaïde Genay,
The University of Melbourne, Australia
Amber Maimon,
Ben-Gurion University of the Negev, Israel

*CORRESPONDENCE
Peter Kullmann,
peter.kullmann@uni-wuerzburg.de

RECEIVED 15 March 2025
ACCEPTED 30 June 2025
PUBLISHED 11 August 2025

CITATION
Kullmann P, Schell T, Botsch M and Latoschik ME (2025) Eye-to-eye or face-to-face? Face and head substitution for co-located augmented reality.
Front. Virtual Real. 6:1594350.
doi: 10.3389/frvir.2025.1594350

Eye-to-eye or face-to-face? Face and head substitution for co-located augmented reality

Peter Kullmann[1]*, Theresa Schell[1], Mario Botsch[2] and Marc Erich Latoschik[1]

[1]Human-Computer Interaction Group, Julius-Maximilians-Universität Würzburg (JMU), Würzburg, Germany, [2]Computer Graphics Group, TU Dortmund University, Dortmund, Germany

In co-located extended reality (XR) experiences, headsets occlude their wearers' facial expressions, impeding natural conversation. We introduce two techniques to mitigate this using off-the-shelf hardware: compositing a view of a personalized avatar behind the visor ("see-through visor") and reducing the headset's visibility and showing the avatar's head ("head substitution"). We evaluated them in a repeated-measures dyadic study (N = 25) that indicated promising effects. Collaboration with a confederate with our techniques, compared to a no-avatar baseline, resulted in quicker consensus in a judgment task and enhanced perceived mutual understanding. However, the avatar was also rated and commented on as uncanny, though participant comments indicate tolerance for avatar uncanniness since they restore gaze utility. Furthermore, performance in an executive task deteriorated in the presence of our techniques, indicating that our implementation drew participants' attention to their partner's avatar and away from the task. We suggest giving users agency over how these techniques are applied and recommend using the same representation across interaction partners to avoid power imbalances.

KEYWORDS

co-presence, diminished reality, mixed reality, augmented reality, virtual reality, avatars

1 Introduction

XR head-mounted displays (HMDs) present an intriguing paradox in co-located social interaction. These devices are inherently personal, displaying visual content mere centimeters from the user's eyes. They can also foster shared experiences through networked software. Integrating the physical environment via pass-through or optical see-through technologies establishes a common reference frame, allowing users to point at and discuss both real or virtual and 2D or 3D objects. However, as noted by Gugenheimer et al. (2019), the very hardware that enables these shared XR experiences physically impedes natural face-to-face communication by obscuring the upper face, arguably even more than when having a conversation with someone wearing sunglasses. This creates an experience where users are physically present and share the same space, yet their ability to convey and interpret subtle nonverbal cues is compromised by the technology intended to enhance their connection. Billinghurst et al. (2003) found that users perceived HMDs as more useful for collaboration if the displays were held on a handle against the face rather than strapped to their heads because it facilitated quick switching between face-to-face conversation and

glances at virtual content. However, doing so could diminish the use of mediated access to a shared virtual world, disrupting the virtual context of the peer interaction.

We perceive faces differently when (partially) occluded. Tinted eyewear makes it harder to identify people and can alter how we attribute social traits such as authority and trustworthiness to the person wearing them (Bartolini et al., 1988; Graham and Ritchie, 2019). Previous work has extensively investigated nonverbal cues in the periocular region, i.e., around the eyes—most notably gaze behavior—and linked them to psychological processes such as turn-taking, attention cueing, and back-channeling (Duncan, 1972; Argyle et al., 1974; Kleinke, 1986; Frischen et al., 2007). The movement of the eyelids, eyebrows, and pupil dilation has been associated with emotional display and other cognitive processes (Ekman and Friesen, 1978; Kröger et al., 2020; Boucher and Ekman, 1975; Hömke et al., 2018).

With progress in HMD technology and advances in 3D reconstruction and animation of human bodies, avatars—humans' digital representation, steered by motion input of who they are representing (Bailenson and Blascovich, 2004)—have become more pervasive, both in experimental laboratory settings and in the wild (Bente et al., 2008; Nowak et al., 2018; Bartl et al., 2021; Menzel et al., 2025), allowing remote headset wearers to feel like they are "being there together." Similarly, some approaches have aimed to inform co-located persons about headset wearers' state, reducing the isolating nature of HMDs as personal viewing devices (Ive et al., 2024; Mai et al., 2017; Matsuda et al., 2021a; Combe et al., 2024). However, when multiple co-located headset wearers participate in a shared extended reality (XR) environment, they cannot see the entire faces of their interaction partners.

We propose using a personalized virtual human model, real-time facial expression tracking, registered device tracking spaces, and blending the avatar with the pass-through view to show co-located XR headset wearers their interaction partner's face, either as a cutout (as if behind a transparent visor) or an avatar head overlay without a headset. Although similar approaches have benefited telepresence scenarios and asymmetric co-located use cases, this is, to our knowledge, the first effort to combine these technologies to mitigate face occlusion by XR headsets in co-located settings. Avatar-mediated face visualization has shown promise in restoring nonverbal communication in co-located XR, even when the digital representation is shown as non-personalized overlay over the pass-through video feed (Combe et al., 2024). However, a direct overlay can create representational conflicts when virtual and physical faces diverge. Prior work on object manipulation in augmented reality has demonstrated that although ownership over a virtual arm is possible, even when their real arm is still visible, some users found the discrepancy between virtual and physical arms distracting (Feuchtner and Müller, 2017). As such, simple compositing approaches suffer from depth-ordering artifacts—where the virtual face may appear incorrectly layered relative to the physical headset. These considerations motivate exploring approaches that better integrate virtual facial representations while maintaining spatial and temporal coherence with the physical interaction context.

The main contributions of this paper are as follows: (1) two novel methods for blending the video pass-through view of an HMD-wearing person with their personalized avatar to recover the face area otherwise occluded by the headset, along with extended implementation details for our shared social XR software using commodity hardware; and (2) a systematic evaluation of these face occlusion mitigation techniques through a within-subject dyadic user study, with recommendations for deploying these techniques in practice.

2 Related work

Our approach is closely related to prior work investigating tracking and rendering human faces when wearing HMDs. In this section, we highlight relevant prior work on virtually recreating HMD-wearing humans and co-located XR interaction from human–computer interaction design, computer vision, and computer graphics.

2.1 XR experience plausibility and congruence

We follow the conceptualization of plausibility as a key construct for XR experiences, as proposed by Latoschik and Wienrich (2022). They model plausibility as emerging from a function of (in) congruence, i.e., (mis-)match between expectations and information processed in cognitive, perceptual, and sensory layers. Central to this congruence-and-plausibility (CaP) model, plausibility then determines XR experience perceptions. Stimuli that are created with sufficient technological development contribute to high plausibility and, consequently, to XR-related qualia. Thus, the perception of an XR experience as plausible depends on how well processed cues match anticipated cues. Since faces are central to face-to-face communication, they have also been central to investigations into computer-mediated communication. Facial expressions are more salient than other bodily regions (Oh Kruzic et al., 2020). Apparatuses that depict facial cues in a manner consistent with face-to-face interactions are beneficial to communication outcomes, e.g., by providing participants with mutual gaze awareness (Ishii et al., 1993; He et al., 2020). We seem to anticipate nonverbal cues present in face-to-face conversation; therefore, recovering access to these cues might improve XR qualia. In this work, we specifically investigate social presence and uncanniness as aspects of qualia.

2.2 Wearable face displays

Prior research has recognized the social challenges posed by HMDs in co-located settings, specifically their tendency to isolate wearers from bystanders. Such use cases are typically called asymmetric since interaction possibilities and/or the presented visual information differ between users (Ens et al., 2019). To address this limitation, researchers have explored various approaches that externalize the HMD wearer's experience with front-facing displays mounted on HMDs, thus lowering the barrier to natural communication. Using "FrontFace," Chan and Minamizawa (2017) presented an abstract visualization of tracked gaze with a pair of cartoon eyes displayed on a front-facing screen.

Shortly after, Mai et al. (2017) introduced "TransparentHMD," a screen mounted on the HMD visor that shows a 3D face model. Notably, the rendered view was perspective-corrected to a single bystander, but the animation (gaze, brows, and lids) was synthesized and shown on a non-personalized 3D face model. In a later study, Mai et al. (2019) compared the TransparentHMD system to two alternatives: visualizing the HMD wearer's state textually on the same front-facing display and a blank-screen baseline. While the collaborative task performance improved more with text display, their avatar representation had positive effects on social presence. Bozgeyikli and Gomes (2022) designed "Googly Eyes," which animated cartoon eyes based on an eye tracker's gaze direction and open/closed states of eyelids. In their evaluation, non-HMD participants collaborating with an HMD participant equipped with Googly Eyes rated the interaction as easier compared to the control group (Bozgeyikli et al., 2024).

Showing a perspective-corrected view is important, but, as argued by Matsuda et al. (2021b), reprojection onto front-facing 2D displays fundamentally suffers from incorrect depth cues. Hence, they advocate for displaying the obscured face region on an outside-facing autostereoscopic display, as used in their "Reverse Pass-Through VR" prototype. Similarly, Ive et al. (2024) proposed an outward-facing 3D display system that shows the headset wearer's face and state, a concept used in the Apple Vision Pro headset as "EyeSight"[1].

The aforementioned approaches have successfully reduced the barrier between HMD wearers and others but are significantly limited by display clarity, resolution, and visor reflections—factors that heavily reduce their utility. In symmetric use cases, where all interaction partners wear headsets, these physical display limitations could be circumvented by emulating a face display in software. However, the efficacy of such an emulated face display would depend heavily on the quality of facial re-construction. Our work resonates well with this line of research while aiming to overcome the shortcomings of physical displays.

2.3 XR face re-enactment

Several previous works have examined self-reenactment—digitally recreating a person' own appearance and behavior—for compositing a rendered face with a camera view of the HMD wearer. Burgos-Artizzu et al. (2015) were likely the first to reconstruct the full view of an HMD wearer's face. Although their method was impressive, even for large occlusions caused by the bulkier HMD visors of that time, estimating upper-face expressions solely from landmarks of the lower face was (and still is) highly limited. Subsequent works used additional camera sensors integrated into the HMDs. Frueh et al. (2017) generated a mixed-reality view of a VR user by capturing them in front of a green screen, placing the virtual environment around them, and rendering a translucent face proxy on the headset region. Later, Thies et al. (2018) presented stereoscopic face views in their "FaceVR" system. They combined the RGB-D camera's view of an HMD wearer and an HMD-internal camera with footage of a target actor captured by a stereo camera rig, thus limiting the output to the target actor's head motion.

In another line of research, behavioral and visual fidelity for telepresence scenarios has seen great advances. For example, Wei et al. (2019) showed impressively realistic reconstructions. Their methods, however, require a specialized capture rig and extensive offline computation upfront, though adding new identities later can require less effort (Cao et al., 2022). A similar work by Ladwig et al. (2024) targeted the use of low-cost commodity hardware, promising commoditization of such approaches.

Now that off-the-shelf AR is becoming more widespread, we expect advances in avatar technology and self-reenactment methods to jointly improve co-located XR interaction. In our work, we aim to build on recent progress in accessible virtual human reconstruction, enabling headset wearers to reenact themselves by tracking facial expressions in real-time.

2.4 Co-located XR embodiment

A prominent advantage of co-located XR is having a familiar, shared reference frame. Since face occlusion fundamentally limits this reference frame's efficacy, some studies have proposed abstract visualizations of interaction partners' occluded state. Piumsomboon et al. (2019) showed co-located users their interaction partner's camera frustum, head direction, or gaze direction. Participants preferred the presence of such awareness cues over a cue-less baseline, and the cues improved performance in a collaborative task. Similarly, Jing et al. (2021) visualized gaze as gaze rays, gaze cursors, or gaze point trails. Participants in their study reported these visualizations to facilitate shared attention and intention and increase the use of deictic references.

Other researchers have focused on recovering the view of a co-located user's face. Takemura and Ohta (2002); (2005) presented a seminal apparatus that used tracked gaze direction and eyelid opening to show a virtual face on top of the other HMD wearer's pass-through footage. Their impressive early work consequently diminished the bespoke hardware system components that were occluded by the virtual (inner) face (Mori et al., 2017). More recent work has used off-the-shelf XR HMDs equipped with face trackers to visualize co-located users as cartoon-like or realistic heads (Combe et al., 2024) but without using a personalized avatar model and without diminishing the headset's appearance.

Another closely related work is "Holoportation" by Orts-Escolano et al. (2016), which transmits 3D reconstructions of objects and people to a remote location. In their qualitative study, users noted that headsets hindered direct eye contact. In response, the authors projected two camera streams filming the eye regions onto a static face mesh model. They stressed the challenges of achieving realism and reported that their system was "not fully over the uncanny valley," suggesting that further work is needed to extend the experience. Our work drives a personalized avatar model and is conceptually closest to that of Combe et al. (2024) and Takemura and Ohta (2002).

2.5 Research objective

In summary, prior work has mostly focused on asymmetric co-located interaction or symmetric telepresence. Their contributions

1 https://support.apple.com/en-us/120481

are helping remote users become closer and bystanders without HMDs become more connected to HMD wearers. Similarly, we suggest that co-located people can recover some of the closeness of face-to-face encounters, targeting the concept of social presence. We explore whether implementing our mitigation techniques is sufficient, given the current maturity of affordable reconstruction methods and face-tracking AR headsets. The decision to use off-the-shelf systems will likely introduce more salient forms of incongruence, such as latency, spatial misalignment, or color mismatch. Although these potential breaks in plausibility might affect the experience, they might still be sufficient to improve the qualia evoked by the XR experience. Hence, we examine social presence as a quale to assess the potential of mitigating face occlusion. To assess how well these techniques work within the inherent limits of technical feasibility, we also investigate perceived uncanniness.

3 Concept

3.1 Virtual human reconstruction

We reconstruct avatars with the smartphone-based pipeline from Menzel et al. (2025). It guides you through capturing multi-view images of a human in two orbits around them. In the body orbit, you take full-shot images of the scanned person standing in A-pose. In the other orbit, you take head close-ups. The image sets are processed further on a dedicated server by segmenting the subject from the background, calculating a 3D point cloud, landmark detection, then template mesh fitting, and finally texturing. The output skinned mesh contains circa 24,000 vertices and 52 facial expression blend shapes, based on Apple ARKit's set of facial expressions[2].

3.2 Avatar animation

We utilize facial expression data provided by Meta Quest Pro's built-in tracking sensors. These are mapped to our mesh model's eye meshes and respective blend shapes. They consist of fourteen expressions for eyelid movement, six for eyebrows, two for nose, eight for cheeks, ten for upper lip, ten for mouth, nine for lower lips, and four for jaw movement.

3.3 Avatar blending

We first register the co-located headset coordinate systems and calibrate how the HMD is placed on the head. Then, we use stenciling to composite the aligned avatar and, for head substitution, the background behind the HMD.

3.3.1 Co-located user coordinate registration

To align the users' local coordinate systems, we use each headset's tracked controllers to sample two points located in opposing room corners. We then move the local coordinate systems' origin to the samples' centroid. We center user interaction around this centroid as alignment accuracy is highest there. As proposed by McGill et al. (2020), this approach is platform-independent and does not require sending data to third-party servers. Alignment quality is determined by tracking error when positioning controllers at the probing landmarks. Thus, we record the controller's pose while it is resting in a rigid controller stand. Some controllers fit snugly into their packaging inserts without disrupting tracking, making them suitable as stands. If packaging inserts are unsuitable and 3D-printing is viable, a controller mount can also be 3D-printed following the method described by Kern et al. (2021b). As button presses on the sampled controller would disturb its position (Wolf et al., 2020), we trigger recording the controller pose via button presses on a second, non-probed controller. Moreover, we filter the raw pose using a one Euro filter (Casiez et al., 2012).

3.3.2 Avatar head registration

Whenever a user puts on the HMD, its placement with respect to the wearer's head is marginally different. This can also change during a session, either by touching the headset or making extensive facial expressions. To align the virtual with the physical head, we need to account for this placement variance. Therefore, we calibrate the headset fit by probing a sparse set of cephalometric landmarks with a 6-DoF controller. Since the upper face is mostly inaccessible for probing, we settled for six landmarks: pronasale (tip of nose), right/left zygion (most protruding point on cheekbones), right/left gonion (mandibular corner), and gnathion (chin). We use the probing point from a virtual controller model as provided by its manufacturer, i.e., the Meta Touch Pro controller's stylus tip. Alternatively, a probing tip can be retrieved following the method proposed by Kern et al. (2021a). We use a set of vertices on the virtual face corresponding to the anatomical landmarks for point-set registration using the rigid Kabsch algorithm (Kabsch, 1976).

3.3.3 Avatar stenciling

To blend the aligned avatar model into the interaction partners' view, we employ stencil testing with an aligned mesh model of the HMD (*cf.* Figure 1). In the absence of a manufacturer-provided model, we propose to use an artist-made model. The pivot of the mesh model might differ from the pivot provided by the device API—in our implementation, this is provided via OpenXR. To correct for this rigid difference, we add a manual positional and rotational offset. To help find a good offset, we probed the visor surface with a self-tracked controller at runtime and adjusted the offset according to how well the virtual controller touches the virtual visor. Alternatively, if no mesh model is available, we propose to use a controller tip, following the method described by Kern et al. (2021a), to record a sparse sample of surface points. Reconstructing a mesh from these points is beyond the scope of this article. In brief, we suggest finding convex sub-regions, generating their convex hull, and merging sub-meshes.

For the see-through visor, we separate the visor area from the mesh model and use it to write to the stencil buffer, masking where to render the avatar face. Following the proposal by Frueh et al. (2017), we pursue "a user experience that conveys a 'scuba mask

2 https://developer.apple.com/documentation/arkit/arfaceanchor/blendshapelocation

FIGURE 1
Co-located headset wearer viewed by another headset wearer **(top row)** is blended with their aligned avatar, either via our "see-through visor" **(center row)** or as "head substitution" **(bottom row)**. See Supplementary Video for demonstrations of all conditions in varied poses.

effect'." We do so by adding a facial gasket mesh. When piloting this gasket visualization, users reacted positively, noting that it made the virtually translucent visor appear more sensible.

For head substitution, we diminish the headset by writing to the stencil buffer over the entire HMD model, thereby revealing a digital twin of the room behind it. In simple use cases of interaction in front of unicolored, evenly lit walls, an unlit plane mesh suffices as a digital twin. We chose this simplified approach for our evaluation study as our controlled laboratory setup, with constant artificial lighting, made a uniform background appropriate. Although more sophisticated photogrammetry workflows exist, they fall outside the scope of our core contributions. Moreover, we apply an alpha texture to hide the body mesh outside the head area. To prevent anyone from seeing the face of the mesh head through the virtual neck, we add an occluder mesh as a seal.

3.4 Software and hardware

We implemented our prototype in Unity v6.0.29 using the Universal Render Pipeline. We used Meta XR Core SDK v71 and Meta Interaction SDK v71 for device input, including facial expression weights, and 3D user interactions, Ubiq v1.0.0-pre.9 for networking (Friston et al., 2021), and Johnathon Selstad's MathUtilities[3] for point set calibration.

We execute the application standalone on a Meta Quest Pro headset (1832×1920 px per eye, 72 Hz target refresh rate) running Meta Quest OS v71. As reported by Wei et al. (2023), its eye tracker has an average accuracy of 1.652° with a precision of 0.699°. A second instance of our application is run on a standalone Meta Quest 3, using Meta Quest OS v72. For networking, we host a custom Ubiq server instance on a dedicated virtual machine in our institutional data center.

4 User study

We aim to gain insights into issues related to recovering the view of a co-located headset wearer. In this section, we describe our two dyadic tasks, motivate our dependent variables, explain the experimental procedure, and describe our sampled population.

We narrow our research interest to two questions:

3 https://github.com/zalo/MathUtilities/

(RQ1) How does mitigating face occlusion influence co-located interaction?

(RQ2) How is an interaction partner perceived when such techniques are applied?

In part, the first aspect is commonly conceptualized as social presence, the "sense of being with another" (Biocca et al., 2003). Research on computer-mediated communication theorized that a medium's richness and users' adaptation to it influence social presence during its use (Oh et al., 2018). Additionally, we select tasks with objective performance metrics. This allows us to determine differences in the collaboration output.

We primarily explore the second question by addressing the constructs of uncanniness. As digital representations of humans approach human-likeness without fully achieving it, they may evoke unease or discomfort, a theory referred to as the uncanny valley (Mori et al., 2012). Moreover, we inquire about participants' overall condition judgments and open feedback. This helps discuss broader, qualitative implications.

4.1 Hypotheses

Overall, we expect our mitigation techniques to impact interaction positively. We assume the following effects:

H1. We expect the display of an avatar (either as a see-through visor or via head substitution) to increase social presence because it uncovers occluded nonverbal behaviors.

H2. We expect the see-through visor to be perceived as more uncanny than head substitution based on two observations: First, any latency-induced misalignment between the avatar and pass-through video feed is likely more salient in the see-through visor condition, where the avatar face is tightly framed within the contours of the real head. Our prototype shows the avatar with a slight delay to the video pass-through, potentially creating an unsettling sense of facial features shifting relative to the skull. In contrast, the transition from avatar head to thinner pass-through neck may act as a perceptual buffer, making misalignment less noticeable. Second, the see-through visor mixes two different sources (avatar and video feed) within the same facial region. This might accentuate dissimilarities between them and result in a less coherent appearance compared to a visually unified face presented in the head substitution condition.

H3. We hypothesize that dyads will perform better in the two avatar conditions compared to the non-avatar baseline, i.e., have more/quicker tower attempts and more completions in the tower task and briefer time to reach a consensus in the image-pairing task. We ground this expectation in the assumption that including avatars restores access to nonverbal cues that are otherwise occluded by the headset, thus facilitating effective interaction. Specifically, access to the interaction partner's eyes enables participants to engage in gaze-based attention-cueing or turn-yielding. These communicative affordances likely support better task performance and more quickly lead to a consensus.

4.2 Tasks

We picked two tasks presenting potential shared experiences in co-located XR. To ensure consistency in interactions and keep preparation manageable, we had participants interact with a confederate. Hence, we considered tasks that are easily repeatable and require neither prior knowledge nor revealing personal information.

4.2.1 Brick tower building

In this executive task, two people work together to recreate a pictured building by stacking 10 colored bricks. Our recreation of the parlor game *Make 'n' Break* has one person act as an architect who sees the target building and verbally describes which stone to put where. The other person acts as a builder, stacking the bricks without a view of the target solution (see Figure 2, left). As soon as the architect recognizes the tower as complete, they can advance to a new, randomly selected target tower, simultaneously returning the bricks to their resting positions along the outer edge of the task area.

4.2.2 Pet-owner portrait pairing

In this judgment task, we let interaction partners pair portrait pictures of dogs with portraits of dog owners. We adapted this task from Hauber et al. (2006), who proposed it with intentionally "highly ambiguous content" to encourage extensive back-and-forth communication between interaction partners. We sourced 65 portrait images from a picture book documenting dog–owner pairs (Schwabe and Vogt, 2014), five of which were picked at random for each trial. Interaction partners are shown a board with five human portraits and dog portraits scattered on the table. Participants are asked to collaboratively find a solution by pinning dog portraits on the board to form pet–owner pairs (see Figure 2, right).

4.3 Measures

To comprehensively evaluate the dynamics between interaction partners and how the blended avatar conditions are perceived, we combined retrospective self-report questionnaires on uncanniness and social presence with performance metrics, preference ratings, and free-text feedback.

4.3.1 Subjective measures

We measured social presence with the Networked Minds Social Presence Inventory (NMSPI) (Biocca and Harms, 2002). Each item has two versions, one targeting the participant's own perception and the other targeting their interaction partner. Responses are marked on 7-point Likert scales (fully agree–fully disagree). They target co-presence (four items for oneself and the other), perceived attentional engagement (three items each), perceived emotional contagion (four items each), perceived comprehension (three each), and perceived behavioral interdependence (three each). We used the NMSPI questionnaire in its entirety (Biocca and Harms, 2003), allowing exploration opportunities for secondary analyses beyond our initially hypothesized effects and enhancing comparability with existing literature. As first-order social presence, respective item ratings are aggregated. As a second-order dimension of social

FIGURE 2
Task interactions. **(Left)** Confederate instructing in the brick tower building task (see-through visor condition). **(Right)** Confederate holding a dog card in the pet–owner portrait pairing task (head substitution condition).

presence, aggregated items are conceptualized as a psycho-behavioral interaction (Biocca and Harms, 2003). Since participants interacted with a confederate and had asymmetric roles in the tower-building task, we disregarded investigating intersubjective symmetry (third-order social presence).

We let participants rate their uncanniness. We measured uncanniness using the uncanny valley index (Ho and MacDorman, 2017). Its semantic differential items are aggregated into the following factors: humanness (five items), attractiveness (four items), and eeriness (nine items), further divided into sub-factors—spine-tingling (five items) and eerie (four items).

To conclude, we asked participants which condition felt closest to face-to-face interaction, most natural, and overall best in a forced-choice task, followed by open-ended prompts for their preference justification and further comments.

4.3.2 Behavioral measures

For the tower-building task, we record the start and stop times of each tower attempt. From these, we derive the number of attempted and/or finished towers and the average tower building time.

For the image pairing task, we measured the consensus time, i.e., the time it took to agree on a solution, from interaction partners' first view of dog and owner portraits to when they signaled content with the pairings.

4.4 Procedure

Our dyadic study took approximately 60 min and was divided into four blocks (*cf.* Figure 3). We obtained written approval for this study from our institutional ethics committee. Each study participant interacted with the experimenter as a confederate. He was familiar with our prototypes but kept blinded to the study condition. Furthermore, he confirmed having no prior acquaintance

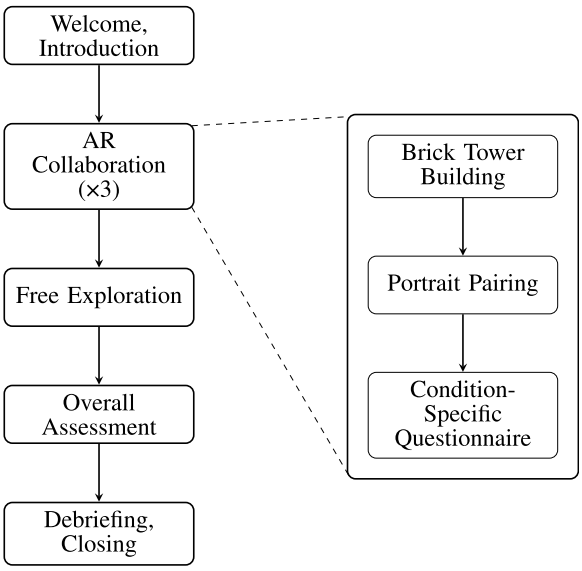

FIGURE 3
Study procedure. Simplified overview; see subsection 4.4 for a detailed description.

with the participants. Using a confederate greatly reduced the apparatus setup and enabled consistent interaction behavior across participants. Although it would have been feasible to reconstruct virtual humans of both interaction partners on the fly, doing so would have added operational complexity and extended the session duration. Since we had prepared the confederate's avatar in advance, we re-used it for all sessions. Device asymmetries further motivated this design choice: the confederate's Meta Quest Pro headset—an older, now-discontinued model—features built-in face-tracking sensors but

offers grainier image quality and lower resolution. In contrast, the participants' Meta Quest 3 headset lacks face-tracking but offers superior video pass-through, including higher-resolution cameras that dynamically adjust exposure, yielding improved color, contrast, and overall visual quality. Throughout the AR exposure, the confederate saw the participant's head occluded by the headset. We instructed him to adopt a pleasant-to-neutral manner, re-initiate conversation during pauses, and gaze at the headset whenever they would spontaneously direct their gaze at their interaction partner.

First, the experimenter welcomed participants and let them read our experiment information, consisting of a briefing, data privacy policy, participation consent, and consent to audio recording. If required, the experimenter answered participants' questions. Then, participants gave informed written consent to their participation and use of their anonymized data. Next, they filled out a demographics questionnaire on a separate laptop.

For the second block, participants sat down at the table with the confederate, and they donned the headsets. The confederate wore the face-tracking Meta Quest Pro HMD, whereas the participants wore a Meta Quest 3 with its superior pass-through image quality. The confederate, hidden from participant view by a virtual wall, instructed them to adjust the HMD straps and lens spacing for optimal display clarity. As a reminder about the procedure, he then informed participants about the upcoming trials, each changing from tasks in XR to subsequent ratings on the questionnaire laptop. Then, the confederate let them familiarize themselves with how to manipulate virtual objects with hand tracking. Subsequent task instructions were provided textually with an accompanying video depicting the main task interactions. Participants could then start the task by pressing a virtual button, thereby revealing the task area and their interaction partner. When the tower task timer expired after 3.5 minutes, the view of the task area and confederate was hidden once more by a virtual wall and instructions for the image pair matching task. As soon as interaction partners agreed upon a solution, the confederate instructed participants to take off the headset and continue with the next questionnaire part. After completing three trials, always with the tasks in the same order, participants donned the HMD once again for a free exploration of the experimental conditions to form an opinion on them. They could switch between conditions by pressing virtual buttons and were encouraged to ask the confederate to perform specific movements, if desired.

Third, participants completed the final sections of the questionnaire, rating the conditions and providing justification for their choices.

Finally, the participants were debriefed and thanked for their participation.

4.5 Participants

We recruited participants in our institution's participation management system. We promoted the study as "Collaborative Puzzles in Augmented Reality" to our pool of students enrolled in bachelor's degree programs in psychology, human–computer interaction, or media communication. For participation eligibility, we required full legal age, no computer addiction, no sensitivity to AR/VR, language fluency, and appropriate corrections for audiovisual impairments. We scheduled sessions during regular working hours and compensated the 26 participants with student credit. We excluded one participant due to insufficient language comprehension. To mitigate potential order effects, we counterbalanced the trial order. Thus, there were four participants per order permutation, except for one trial order ("no avatar," "see-through visor," "head substitution"), which had five.

The 25 included participants were 18–27 years old (M: 21, SD: 2.1), mostly native German speakers (one non-native reported fluency), had normal hearing, normal or corrected-to-normal visual acuity, and normal color vision. Four participants self-identified as male, while the remaining participants identified as female. Nineteen reported playing video games for less than 1 hour a day, and the other six reported 1–3 hours of daily video game usage. Regarding prior AR/VR experience, seven stated to have experienced less than 1 hour, ten experienced 1–3 hours, four reported 3–5 hours of experience, two participants reported 5–10 hours of experience, and two reported more than 20 hours of prior AR/VR experience.

5 Results

We examined differences between conditions across several measures. Below, we present analyses for each measure separately, followed by a summary of the overall pattern of results.

We analyzed data using R Statistical Software v4.4.2 (R Core Team, 2024). For uncanniness, social presence, and task metrics, we employed multilevel modeling with maximum likelihood estimation and random effects (Pinheiro and Bates, 2000; Pinheiro et al., 2024). Planned orthogonal contrasts tested (1) no avatar vs both avatar conditions (see-through avatar and head substitution) and (2) see-through avatar vs head substitution (Hothorn et al., 2008). Note that while we report likelihood ratio tests for overall condition effects, our primary hypotheses are tested through the planned orthogonal contrasts, which may detect specific differences even when the overall test does not reach significance. This approach provides greater statistical power for testing our *a priori* hypotheses regarding the specific pattern of differences between conditions.

Descriptive statistics are presented in Table 1. We report effects as significant at an alpha cutoff of 0.05.

5.1 Uncanniness

We examined uncanniness through its dimensions: humanness, attractiveness, and eeriness (with sub-dimensions, spine-tingling and eerie) (see Figure 4).

Multilevel modeling showed no significant differences between conditions for the dimension *humanness* ($\chi^2(2) = 5.10$, $p = .078$, and $\eta_p^2 = .10$). The two avatar conditions were rated as less human than the no-avatar baseline (planned contrast 1, $b = .25$, $t(48) = 2.25$, and $p = .029$), while the two avatar conditions did not differ significantly (planned contrast 2, $b = .008$, $t(48) = .04$, and $p = .97$).

For *attractiveness*, ratings markedly differed across conditions ($\chi^2(2) = 5.11$, $p = .078$, and $\eta_p^2 = .10$). The no-avatar condition was

10.3389/frvir.2025.1594350

TABLE 1 Statistical results. Descriptive and inferential statistics: overall effect (χ^2 (2), *p*-value, and η_p^2), contrast 1 (comparison between no-avatar baseline and two avatar conditions: regression coefficient b, t-statistic, and *p*-value), and contrast 2 (comparison between two avatar conditions: regression coefficient b, t-statistic, and *p*-value).

Factor	Condition	Mean (SD)	Overall effect	Contrast 1	Contrast 2
Humanness	No avatar	5.44 (1.57)	5.10, 0.078, and 0.10	0.25, 2.25, and **0.029**	0.08, 0.04, and 0.97
	S.-t. visor	4.71 (1.20)			
	Head sub.	4.70 (1.21)			
Attractiveness	No avatar	5.32 (0.98)	5.11, 0.078, and 0.10	0.20, 2.22, and **0.03**	−0.075, −0.49, and 0.62
	S.-t. visor	4.66 (1.26)			
	Head sub.	4.81 (1.23)			
Eeriness (total)	No avatar	3.19 (0.69)	15.09, **0.0005**, and 0.26	−0.29, −4.06, and **0.0002**	−0.36, −0.29, and 0.77
	S.-t. visor	4.02 (0.97)			
	Head sub.	4.09 (0.98)			
Eeriness (eerie)	No avatar	2.93 (1.07)	12.58, **0.0018**, and 0.22	−0.02, −3.64, and **0.0007**	−0.02, −0.12, and 0.91
	S.-t. visor	4.00 (1.26)			
	Head sub.	4.04 (1.34)			
Eeriness (spine-tingling)	No avatar	3.44 (0.67)	15.89, **0.0004**, and 0.27	−0.22, −4.20, and **0.0001**	−0.05, −0.58, and 0.56
	S.-t. visor	4.04 (0.89)			
	Head sub.	4.14 (0.86)			
Co-presence (self)	No avatar	6.18 (1.08)	2.85, 0.24, and 0.06	0.01, 0.17, and 0.86	−0.20, −1.67, and 0.10
	S.-t. visor	5.95 (1.07)			
	Head sub.	6.34 (0.90)			
Co-presence (other)	No avatar	6.16 (0.95)	0.66, 0.72, and 0.01	0.05, 0.73, and 0.47	0.04, 0.32, and 0.75
	S.-t. visor	6.04 (1.09)			
	Head sub.	5.96 (1.04)			
Perceived attentional engagement (self)	No avatar	5.12 (1.15)	3.40, 0.18, and 0.07	−0.08, −1.53, and 0.13	−0.09, −1.03, and 0.31
	S.-t. visor	5.95 (1.07)			
	Head sub.	6.34 (0.90)			
Perceived attentional engagement (other)	No avatar	5.76 (0.93)	1.13, 0.57, and 0.02	−0.04, −0.98, and 0.33	−0.03, −0.38, and 0.71
	S.-t. visor	5.85 (0.89)			
	Head sub.	5.91 (0.82)			
Perceived emotional contagion (self)	No avatar	3.37 (1.30)	1.35, 0.50, and 0.31	−0.06, −1.13, and 0.26	−0.02, −0.21, and 0.84
	S.-t. visor	3.54 (1.35)			
	Head sub.	3.58 (1.42)			
Perceived emotional contagion (other)	No avatar	3.55 (1.37)	2.94, 0.23, and 0.06	−0.06, −1.15, and 0.26	0.11, 1.26, and 0.21
	S.-t. visor	3.82 (1.57)			
	Head sub.	3.61 (1.55)			
Perceived comprehension (self)	No avatar	6.12 (0.92)	0.06, 0.97, and 0.001	−0.002, −0.05, and 0.96	0.02, 0.24, and 0.81
	S.-t. visor	6.15 (0.69)			
	Head sub.	6.11 (0.87)			

(Continued on following page)

TABLE 1 (*Continued*) Statistical results. Descriptive and inferential statistics: overall effect (χ^2 (2), *p*-value, and η_p^2), contrast 1 (comparison between no-avatar baseline and two avatar conditions: regression coefficient b, t-statistic, and *p*-value), and contrast 2 (comparison between two avatar conditions: regression coefficient b, t-statistic, and *p*-value).

Factor	Condition	Mean (SD)	Overall effect	Contrast 1	Contrast 2
Perceived comprehension (other)	No avatar	5.92 (0.96)	1.34, 0.51, and 0.03	−0.05, −1.07, and 0.29	0.03, 0.40, and 0.69
	S.-t. visor	6.11 (0.68)			
	Head sub.	6.04 (0.90)			
Perceived behavioral interdependence (self)	No avatar	5.27 (1.07)	0.68, 0.71, and 0.01	−0.03, −0.56, and 0.58	−0.06, −0.58, and 0.56
	S.-t. visor	5.31 (1.21)			
	Head sub.	5.43 (1.36)			
Perceived behavioral interdependence (other)	No avatar	5.27 (1.28)	10.80, **0.005**, and 0.19	−0.10, −2.95, and **0.005**	−0.03, −0.56, and 0.58
	S.-t. visor	5.47 (1.31)			
	Head sub.	5.66 (1.02)			
Average tower duration	No avatar	50.2 (19.3)	3.72, 0.16, and 0.09	−3.07, −1.61, and 0.12	3.30, −1.61, and 0.12
	S.-t. visor	62.8 (39.9)			
	Head sub.	56.0 (18.5)			
Tower duration (attempts)	No avatar	4.62 (1.43)	7.32, **0.026**, and 0.16	0.19, 2.71, and **0.0099**	0.002, 0.02, and 0.99
	S.-t. visor	4.05 (1.43)			
	Head sub.	4.05 (1.07)			
Tower finishes	No avatar	4.00 (1.58)	11.54, **0.003**, and 0.24	0.24, 3.41, and **0.0016**	−0.096, −0.79, and 0.43
	S.-t. visor	3.19 (1.44)			
	Head sub.	3.38 (1.12)			
Consensus duration	No avatar	142.4 (45.4)	3.16, 0.21, and 0.08	4.33, 1.55, and 0.13	3.88, 0.80, and 0.43
	S.-t. visor	133.5 (38.1)			
	Head sub.	125.2 (40.1)			

Bold type indicates statistically significant values.

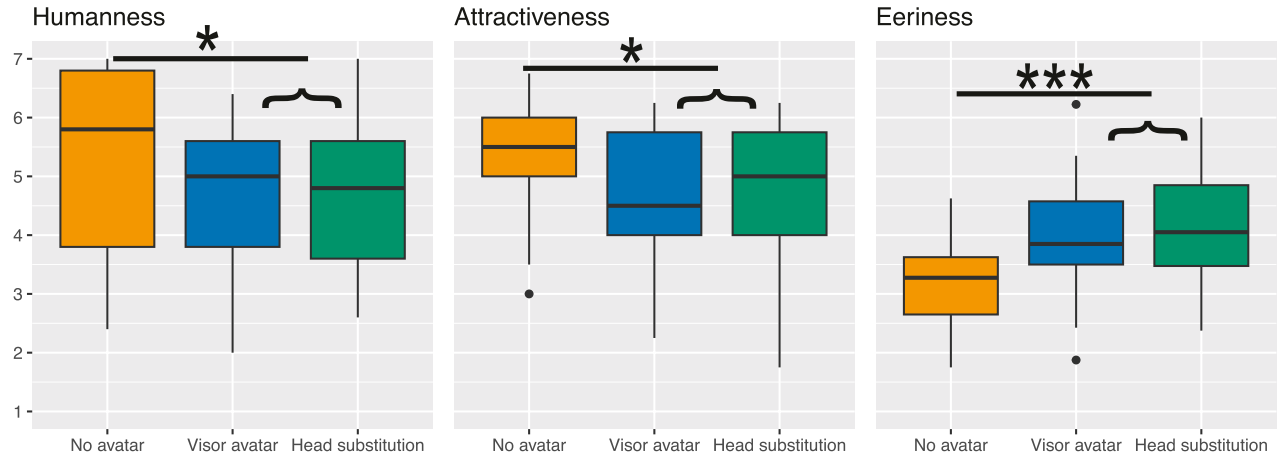

FIGURE 4
Uncanny valley index ratings. Box plots display the median (center line), interquartile range (box hinges), and range of values (whiskers), excluding outliers (points further from the hinge than 1.5 * interquartile range). Single asterisks signify $p < 0.05$, and the triple asterisk signifies $p < 0.0001$.

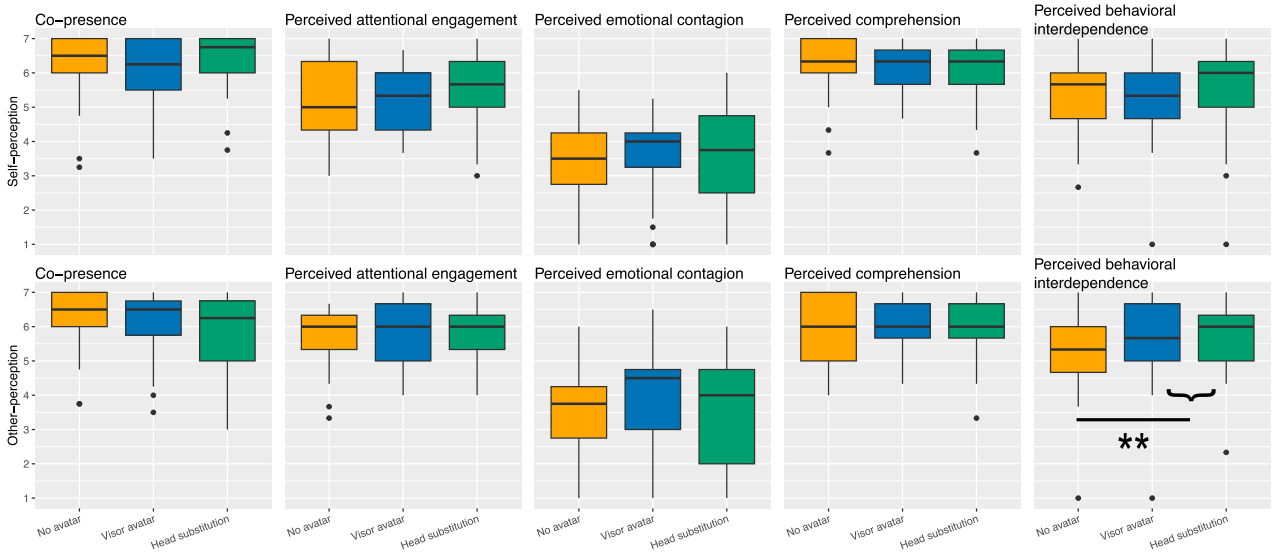

FIGURE 5
Network Minds Social Presence Inventory scores. Box plots display the median (center line), interquartile range (box hinges), and range of values (whiskers), excluding outliers (points further from the hinge than 1.5 * interquartile range). The double asterisk signifies $p < 0.01$.

rated as significantly more attractive than the avatar conditions ($b = .20$, $t\,(48) = 2.22$, and $p = .03$). No significant difference emerged between the two avatar conditions ($b = -.075$, $t\,(48) = -.49$, and $p = .62$).

Eeriness ratings showed a significant overall effect ($\chi^2\,(2) = 15.09$, $p = .0005$, and $\eta_p^2 = .26$), with the no-avatar showing significantly more eeriness than the avatar conditions ($b = -.29$, $t\,(48) = -4.06$, and $p = .0002$), but no significant differences were observed between avatar conditions ($b = -.036$, $t\,(48) = -.29$, and $p = .77$). Both sub-dimensions followed similar patterns. The spine-tingling sub-dimension differed significantly ($\chi^2\,(2) = 15.89$, $p = .0004$, and $\eta_p^2 = .27$), as did the eerie sub-dimension ($\chi^2\,(2) = 12.58$, $p = .0018$, and $\eta_p^2 = .22$). Planned contrasts revealed that the no-avatar condition was less eerie ($b = -.02$, $t\,(48) = -3.64$, and $p = .0007$) and less spine-tingling ($b = -.22$, $t\,(48) = -4.20$, and $p = .0001$) than the avatar conditions, which themselves neither differed in the sub-dimension eerie ($b = -.02$, $t\,(48) = -.12$, and $p = .91$) nor spine-tingling ($b = -.05$, $t\,(48) = -0.58$, and $p = .56$).

Contrary to our hypothesis, these results suggest that participants' uncanniness perceptions did not differ between the two avatar conditions. Notably, both were perceived less favorably than the no-avatar baseline.

5.2 Social presence

Participants' sense of social presence was evaluated in two dimensions: co-presence (first-order social presence) and psycho-behavioral interaction (second-order social presence). Each dimension was assessed from participants' self-perception and their perception of their interaction partner (see Figure 5).

Perceived behavioral interdependence in other-perception differed significantly between conditions ($\chi^2\,(2) = 10.80$, $p = .005$, and $\eta_p^2 = .19$). Ratings were significantly lower without an avatar than in the two avatar conditions ($b = -.1$, $t\,(48) = -2.95$, and $p = .005$) and slightly lower for the see-through avatar condition compared to head substitution ($b = -.03$, $t\,(48) = -.56$, and $p = .58$).

Other social presence factors showed no significant differences between conditions. Specifically, ratings did not differ significantly for co-presence (self-perception: $\chi^2\,(2) = 2.85$, $p = .24$, and $\eta_p^2 = .06$ and other-perception: $\chi^2\,(2) = .66$, $p = .72$, and $\eta_p^2 = .01$), perceived attentional engagement (self: $\chi^2\,(2) = 3.4$, $p = .18$, and $\eta_p^2 = .07$ and other: $\chi^2\,(2) = 1.13$, $p = .57$, and $\eta_p^2 = .02$), perceived emotional contagion (self: $\chi^2\,(2) = 1.35$, $p = .50$, and $\eta_p^2 = .03$ and other: $\chi^2\,(2) = 2.94$, $p = .23$, and $\eta_p^2 = .06$), and perceived comprehension (self: $\chi^2\,(2) = .06$, $p = .97$, and $\eta_p^2 = .001$ and other: $\chi^2\,(2) = 1.34$, $p = .51$, and $\eta_p^2 = .03$) and self-perception of participants' behavioral interdependence ($\chi^2\,(2) = .68$, $p = .71$, and $\eta_p^2 = .01$).

Overall, the view of an avatar, compared to seeing an interaction partner occluded by their headset, fostered social presence in the aspect of mutual understanding with one's interaction partner.

5.3 Task metrics

We annotated voice recordings of the tasks with ELAN (Brugman et al., 2004)[4] by marking timestamps of task sounds emitted as feedback from virtual button pokes. Since four recordings were (partially) corrupt, we considered the remaining 21 complete annotations.

4 Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands, retrieved from https://archive.mpi.nl/tla/elan

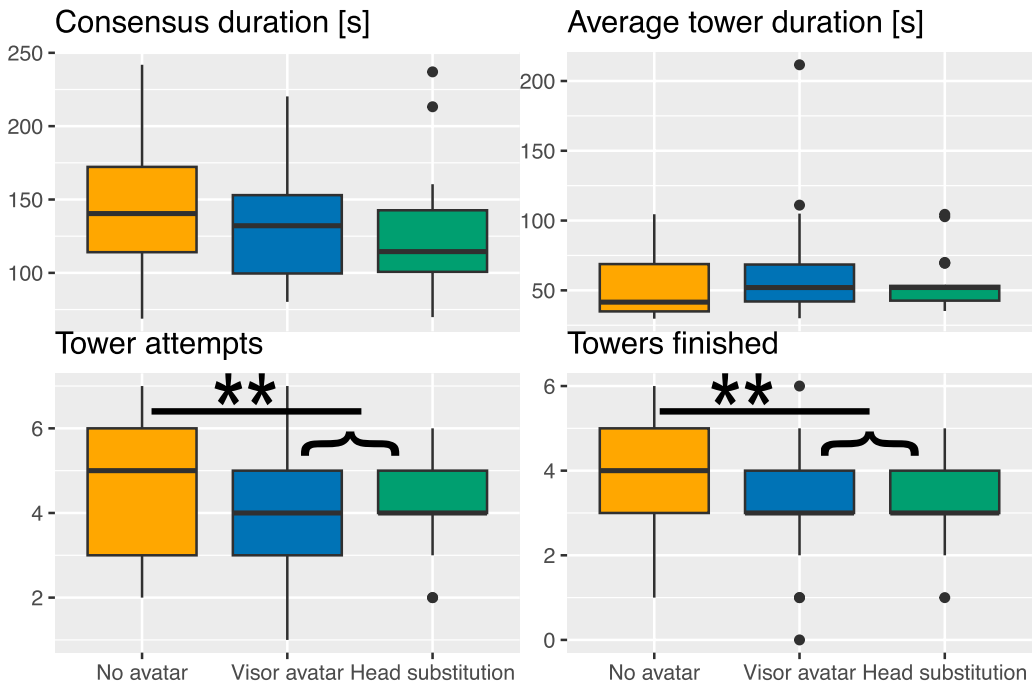

FIGURE 6
Task metrics. Box plots display the median (center line), interquartile range (box hinges), and range of values (whiskers), excluding outliers (points further from the hinge than 1.5 * interquartile range). Double asterisks signify $p < 0.01$.

5.3.1 Tower-building performance

We analyzed the performance in the tower-building task by assessing average tower-building duration, number of build attempts, and number of build completions across conditions (see Figure 6).

Preliminary analysis revealed a notable practice effect across trials, particularly after the first trial, prompting us to include trial number as a covariate in our analyses. After controlling for practice effects, we found no significant effects between conditions on average build duration ($\chi^2(2) = 3.72$, $p = .16$, $\eta^2_{pCondition} = .09$, and $\eta^2_{pTrial} = .29$). However, significant differences emerged in both the number of build attempts ($\chi^2(2) = 7.32$, $p = .03$, $\eta^2_{pCondition} = .16$, and $\eta^2_{pTrial} = .41$) and build completions ($\chi^2(2) = 11.54$, $p = .003$, $\eta^2_{pCondition} = .24$, and $\eta^2_{pTrial} = .37$).

Planned orthogonal contrasts revealed that both avatar conditions led to lower numbers of build attempts compared to the baseline condition ($b = .19$, $t(38) = 2.71$, and $p = .01$) and build completions ($b = .23$, $t(38)3.40 =$, and $p = .002$). The two avatar conditions did not differ significantly from each other in either attempts ($b = .002$, $t(38) = .02$, and $p = .99$) or completions ($b = -.10$, $t(38) = -.79$, and $p = .43$). These results were generally robust to the inclusion of trial number as a covariate, with the exception of build attempts, which became significant only after controlling for practice effects.

In summary, the display of an avatar had a negative impact on tower-building performance. The effect was similar for both avatar conditions.

5.3.2 Consensus time

For each trial, we measured the time participants spent to reach consensus on image pairings (see Figure 6).

Descriptive statistics suggested faster consensus in both avatar conditions (see-through visor: $M = 1133.5$s and $SD = 38.1$s; head substitution: $M = 125.2$s and $SD = 40.1$s) compared to the no-avatar baseline ($M = 142.4$s and $SD = 45.4$s), but inferential statistics did not support this pattern. Conditions had no overall significant effect on consensus time ($\chi^2(2) = 3.2$, $p = .20$, and $\eta^2_p = .07$). Neither the planned contrast between baseline and avatar conditions showed significant differences ($b = 4.33$, $t(40) = 1.56$, and $p = .13$) nor did the contrast between the two avatar conditions ($b = 4.18$, $t(40) = .87$, and $p = .39$).

Although these results trend in our expected direction, suggesting potential benefits of blending in an avatar for co-located collaboration, the high variability in consensus times might have contributed to the lack of statistical significance.

5.4 Condition preferences

Participants indicated which condition they considered most similar to face-to-face interaction, most natural, and overall best. For each dimension, they picked one condition as their favorite. We analyzed it with Pearson's chi-squared test for count data and report Cramer's V measure of effect size (Navarro, 2015).

When asked which condition felt closest to the face-to-face interaction, there was no significant difference between conditions ($\chi^2(2) = 3.92$, $p = .14$, and Cramer's $V = .28$), with head substitution receiving 12 votes (48%), the no-avatar condition 9 votes (36%), and the see-through visor 4 votes (16%).

The choice of the most natural condition differed significantly from a uniform distribution ($\chi^2(2) = 13.76$, $p = .001$, and Cramer's

$V = .52$), with a strong preference for the no-avatar condition (17 votes, or 68%) over the see-through visor (5 votes, 20%) and head substitution (3 votes, 12%).

For the overall preferred condition, differences were not significant ($\chi^2(2) = 5.84$, $p = .05$, and Cramer's $V = .34$), with 14 participants favoring the no-avatar condition (56%), whereas see-through visor and head substitution received 5 (20%) and 6 votes (24%), respectively.

5.5 Qualitative feedback

We used reflexive thematic analysis (Braun and Clarke, 2006; Braun and Clarke, 2019) to group how participants justified their condition preferences. We derived themes through iterative coding and reflexive engagement with the data. Many participants highlighted having had fun, and several enjoyed the tasks, though two considered tower-building "tricky," and another one disliked the image pairing task as it left "much room for interpretation." We clustered responses relevant to the different representations of their interaction partner into three key themes that characterize participant experiences: importance of gaze cues, artificiality and uncanniness, and truthfulness to the apparatus.

5.5.1 Importance of gaze cues

Gaze cues were consistently commented on as critical for natural and effective interaction. Participants consistently emphasized the importance of eye contact and gaze direction for meaningful interaction. For example, one participant stated that tracked eyes enabled them to know "whether the other person is looking at me and whether I need to react or listen, *etc.*,"" indicating the value of this feature for social interaction.

Conversely, the absence of gaze cues in no-avatar condition was identified as problematic; participants reported the view of the face-occluding HMD with no avatar to be "unsettling" and "robot-like" and noted that it made them "uncertain whether the person is looking at me when making a statement. It also felt a bit distant."

5.5.2 Artificiality and uncanniness

Participants frequently commented on the artificial nature of the avatars and how their appearance or behavior detracted from the experience.

Numerous participants mentioned an overall discomfort with the avatar conditions, calling them "unnatural," "unpleasantly artificial," and "evidently artificial and therefore insincere and not real." While most participants mainly expressed a general unease with the depictions or that they "lacked realism," one explicitly explained their perception of the avatar conditions as unnatural "because the eyes were wide open and the facial expressions were stiff and unnatural." The head substitution condition was praised for providing the "most legible emotions and facial expressions." However, others judged the avatar's mouth region as "distracting" or "spooky." Critiques stating that "mouth movement is well visible without avatar" highlight how flawed lower-face representation in the avatar eroded the quality of interaction.

Some noted shortcomings in how avatars were integrated into the immersive environment. For head substitution, one criticism was that "the background behind the head was brighter than the rest of the background." For the see-through avatar, one participant expressed it to "seem strange, probably because of the outlined eyes," referring to the virtual face gasket. These issues detracted from the seamlessness of the experience, apparently diminishing the overall acceptance of the avatars.

5.5.3 Truthfulness to the apparatus

The third major theme reflects participants' expectations conflicting between the ideal of face-to-face interaction and the physical reality of wearing an HMD. Participants often grappled with the tension between avatars mimicking face-to-face interaction ("this was the best way to feel in contact with the other person") and acknowledging the HMD's physical presence ("The view of the virtual glasses on the head was the most natural because it looked just like you would have seen the person in real life"). This demonstrates different priorities: some valued truthfulness to the physical situation, while others sought idealized face-to-face interaction qualities. One participant justified their preference for the see-through visor condition by having access to eye movements while "one was still aware of being in the virtual world because the glasses are still visible." Hence, the see-through visor seems to have more technical honesty in mediation.

6 Discussion

Results from our repeated-measures evaluation in a dyadic setup show several benefits of our avatar blending techniques.

6.1 Subjective measures

6.1.1 Social presence

In line with hypothesis H1, we observed a slight increase in social presence in either avatar condition compared to the no-avatar baseline. Our analysis revealed that participants' perception remained largely consistent across conditions, with most social presence measures showing no significant differences. Notably, we observed a significant positive shift in one dimension in the presence of our mitigation techniques: participants reported enhanced perceived behavioral interdependence with their interaction partner. This factor captures the extent to which interaction partners feel their actions are dependent on and synchronized with their partner's behavior, indicating more genuine two-way social interaction rather than passive observation. These findings align with several prior studies that have investigated social presence in similar contexts. Mai et al. (2019) found similar social presence across conditions when comparing a front-facing screen, either representing the wearer's state via text, an avatar, or not at all (blank screen), despite using synthetic avatar face animations instead of tracking expressions. Similarly, Bozgeyikli et al. (2024) reported comparable social presence for tablet users when either seeing their co-located interaction partner's tracked gaze represented on an HMD-mounted screen or seeing a blank screen. Still, they noted that tablet users in the gaze representation condition felt more connected to and aware of their partners while expressing greater task

confidence. Our findings of enhanced *perceived behavioral interdependence* in the presence of avatars resonates well with Combe et al. (2024), who also observed an increase in this dimension of social presence when superimposing avatars onto the video pass-through view of interaction partners—albeit without blending. However, theirs was not affected by avatar presence but by the animation type, with audio-driven mouth animation yielding higher ratings than live-tracked animation of the mouth region. This suggests that the specific avatar representation may be critical in enhancing certain aspects of social presence in co-located XR.

6.1.2 Uncanniness

In contrast to our hypothesis H2, the two avatar conditions (see-through visor and head substitution) did not differ significantly from each other in their effect on uncanniness but were rated as more uncanny (less human, significantly less attractive, and significantly more eerie) than the baseline (no avatar). This aligns with prior work demonstrating perceptual challenges of avatars in social augmented reality. Mai et al. (2019) found participants to be either excited or uncertain about an HMD-mounted screen representing its wearer's state, reporting that approximately a third of participants found avatar representations uncanny, although they did not inquire about uncanniness ratings for their blank-screen baseline condition. Although we hypothesized that the see-through visor condition would be perceived as more uncanny than head substitution, data did not support this directional prediction. Retrospective consideration reveals equally valid theoretical arguments for expecting the opposite pattern of a directional effect, suggesting that our hypothesis may have been overly specific. The see-through visor preserves the non-occluded portions of the real face, keeping its realism at the level of the pass-through video feed. Furthermore, it minimizes the extent of the rendered avatar, particularly eliminating the need for the inner mouth region (with teeth and tongue) and hair, which are traditionally challenging to reconstruct at high fidelity. These competing arguments highlight a fundamental gap in how we understand uncanniness in mixed-reality contexts. Previously, Kullmann et al. (2025) found that participants who viewed their mirrored self-avatar with a static mouth region rated it as less uncanny than those who viewed their virtual mouth animated using live tracking data. They attributed higher uncanniness and lower self-identification to insufficient reconstruction and/or animation quality of the oral cavity.

6.1.3 Preference ratings and justifications

The no-avatar baseline condition was significantly more often rated most natural, and head substitution was judged closest to face-to-face interaction by markedly more participants than other conditions. Participants' overall impression ratings (pick of most natural, most similar to face-to-face interaction, and overall best condition) and their justifications thereof emphasize a complex interplay between social interaction quality and technological functionality. Participant feedback indicates that although the avatars were noticeably artificial—even evoking some uncanny responses—they still managed to recover critical nonverbal cues. Commonly, participant expectations conflicted between replicating face-to-face interaction and the physical reality of wearing an HMD. Some appreciated the familiarity of head substitution as a simulation of face-to-face interaction, whereas others preferred the see-through visor since it remained true to the technical apparatus—displaying nonverbal behavior of the eye region while still signaling the presence of the HMD.

6.2 Behavioral measures

For improvements in both task types, as assumed in hypothesis H3, results were mixed. In trials with either avatar technique, compared to the no-avatar baseline, participants took markedly less time in the image pairing task to reach a consensus. This fits prior research on mediated group work. Tasks with more ambiguity typically benefit more from communication media providing more access to subtle nonverbal cues (Straus and McGrath, 1994). Accordingly, the recovered eye region in our techniques reveals its phatic function that helps regulate communication, e.g., taking conversational turns (Surkamp, 2014). Notably, performance in the tower-building task worsened in the presence of either avatar technique. We suspect this to be caused by the novelty of our methods and participants' relatively short prior XR experience. The avatar appeared highly salient to participants and may have distracted them from their actual task in the builder role. As argued by Straus and McGrath (1994), there is less demand for social context in tasks with objectively correct solutions.

6.3 Limitations and future work

Several limitations of our implementation and evaluation should be noted.

First, the avatar head was still offset from the real head after our face landmark calibration, even if no study participants explicitly mentioned this error. Similar to the mismatch between virtual controller mesh models and controllers, as depicted in the pass-through video, this is more apparent in closer proximity to the observer. This is limited by mismatches in camera parameters between the game engine video pass-through camera (Chaurasia et al., 2020), pass-through reprojection artifacts, and the accuracy of the avatar scale. Future work should resolve the mismatch in the avatar scale by increasing the scale of the virtual head or diminishing a buffer zone around the virtual head.

Second, since we employed two different camera systems—a smartphone for scanning the user's body and the headset's built-in pass-through cameras for video-see-through augmented reality—we have to rely on manual configuration of avatar shading parameters. Previously proposed color-matching workflows would allow refining and automating this process (Takemura et al., 2006). Similarly, images for the body scan could also be taken from the headset itself, reducing the need for extensive color matching.

Third, our face animation method could be improved. Capturing person-specific expression manifestations and fitting them to the avatar mesh as personalized facial blend shapes, as proposed by Menzel et al. (2022), would result in more natural avatar expressions.

One simplification in our evaluation was to perform it in a room with fixed lighting, placing the confederate in front of a

monochrome wall. More elaborate backgrounds could be reconstructed either as an offline step or at runtime, as surveyed by Mori et al. (2017), allowing less restricted use. Relevant prior work has explored such more complex approaches: Feuchtner and Müller (2017) employed real-time in-painting to recover the egocentric view of a room occluded by a headset wearer's arms, while Kari et al. (2023) combined pre-scanned digital twins of rooms and objects to enable virtual objects to seemingly affect physical elements.

Longer or repeated exposure to the imperfect avatars might also lessen their uncanny perception, following prior work that indicated slight habituation to increased exposure to a robot (Złotowski et al., 2015) and work that showed that unnatural behavior of avatars can become less noticeable over time (Dobre et al., 2022). Better yet, advancing the avatar's depiction could help resolve the tension between its uncanniness and the usefulness of accessing nonverbal cues.

Moreover, we decided to employ a confederate as the study participants' interaction partner and only apply our techniques to him. Although this approach limits the naturalism of the interaction and may reduce the sense of mutual engagement, it offers several practical and methodological advantages. Most importantly, it allowed for a consistent and standardized explanation style in the tower-building task across participants and minimized the effects of varying avatar attractiveness, thus increasing internal validity. Another practical constraint justified this decision: the confederate's device offered face-tracking but poor pass-through quality, while the participants' device lacked face-tracking but offered higher pass-through resolution. Running truly interactive dyads would have required equipping participants with the same device, markedly degrading their AR experience. Given these considerations, the benefits of using a confederate outweighed its limitations. Future evaluations should consider applying our techniques to both interaction partners and investigate emerging communication patterns such as joint attention and mutual gaze behaviors (Jording et al., 2018). Gaze behavior's fundamentally bidirectional nature could thus facilitate emotional mimicry (Mauersberger et al., 2022). Future work could explore showing an avatar in contexts without tracked facial expressions, such as when HMDs lack sensors or restrict access to tracking data. Prior work has shown that synthetic avatar facial animation can be preferred over static faces, sometimes even over veridical behavior (Borland et al., 2013; Seele et al., 2017; Kullmann et al., 2023). Co-located interaction could be improved by revealing an interaction partner's face with synthesized eye movement (Canales et al., 2023).

In line with uncanniness ratings, the avatar depiction had considerable flaws in appearance and/or behavior. This reinforced perceptions of artificiality, deteriorating their interaction and perception of their interlocutor's depiction. However, several participants tolerated imperfect avatar appearance and behavior since they preserved gaze utility. Further theoretical development and empirical investigation are needed to better understand the complex interplay of factors contributing to a congruent representation of a co-located interaction partner by combining real and virtual facial elements.

Our implementation of the proposed mitigation techniques has the aforementioned limitations, which, in the terminology of the congruence-and-plausibility model (Latoschik and Wienrich, 2022), introduce incongruence across sensory, perceptual, and cognitive layers. Artifacts are visible in Figure 1 and evident in qualitative participant feedback. These limitations are further reflected in participants rating both avatar conditions as more uncanny than the no-avatar baseline. Despite (expected) weaknesses of limited implementation, the evaluation data already showed positive effects of our techniques.

7 Conclusion

Our goal was to investigate how mitigating face occlusion influences co-located interaction and perception of one's interaction partner. We proposed two mitigation techniques: (1) a see-through visor—diminishing the visor front plate to reveal the wearer's avatar, and (2) head substitution—masking out the entire headset and superimposing the wearer's avatar head.

We had assumed that we could improve co-located XR with our avatar techniques. Both techniques markedly improved interaction aspects, although their impact was less pronounced during the executive task in our evaluation and accompanied by reservations regarding their uncanny appearance. The negative impact of avatar conditions on performance metrics could reflect increased cognitive load from the presence of avatars or potentially that the avatars distracted participants from the building task itself. Since both avatar conditions had similar effects, the mere presence of an avatar might have drawn attention, thus deteriorating performance.

For use in future prototypes, we draw comparisons to early work by Argyle et al. (1968). They showed a dual effect of obscuring one interaction partner's eye region with tinted glasses: the wearer tended to feel more dominant and comfortable, while their interlocutor felt less so. Similar to how sunglasses have been described as a "social shield" with lopsided effects (Viola, 2024), having one of two interaction partners wear AR glasses has been shown to elicit imbalanced effects (Chung et al., 2023). Therefore, we advocate for responsibly informing users about their representation and empowering their agency by having them configure it. This could involve showing users a preview of how their avatar will appear to others before activating it for interaction partners—similar to camera previews in video conferencing—or implementing these techniques as opt-in features. Representing co-located headset wearers in the same manner may avoid introducing power imbalances due to uneven access to each other's facial cues.

In summary, we presented a see-through visor and head substitution—two techniques to support co-located headset-based XR interaction. Our combination of personalized virtual human models, facial expression tracking, and controller-based registrations allows us to recover the view of the otherwise occluded face of a co-located XR headset wearer. We presented an evaluation in a dyadic setting and demonstrated its positive effect on social presence and performance in a judgment task. In particular, the mitigation techniques increased participants' perceived mutual understanding with their interaction partner. In our judgment task, we observed descriptively quicker consensus when applying our mitigation techniques. We suggested offering the techniques as a user-driven option. Connecting co-located headset wearers is pertinent to improving co-located XR scenarios. We trust

researchers and practitioners to expand this approach, facilitating similar use cases.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://osf.io/rp3fn/ (DOI: 10.17605/OSF.IO/RP3FN).

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Institute for Human–Computer-Media at Julius-Maximilians-Universität Würzburg. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

PK: Visualization, Project administration, Validation, Formal Analysis, Writing – review and editing, Data curation, Methodology, Conceptualization, Writing – original draft, Software, Investigation. TS: Conceptualization, Software, Writing – review and editing. MB: Funding acquisition, Conceptualization, Resources, Methodology, Validation, Supervision, Writing – review and editing, Software. ML: Project administration, Writing – review and editing, Validation, Methodology, Conceptualization, Resources, Supervision, Funding acquisition.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research has been funded by the Bavarian State Ministry for Digital Affairs in the project XR Hub (Grant A5-3822-2-16) and the German Federal Ministry of Education and Research (BMBF) in the project ViLeArn More (16DHB2214).

Acknowledgments

The authors thank Jinghuai Lin for useful suggestions regarding shader development and reviewers for their helpful comments.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. During the preparation of this manuscript, we utilized Anthropic Inc.'s Claude 3.5 Sonnet, a generative artificial intelligence tool, solely for language editing purposes as non-native English speakers. The scientific content, data analysis, interpretations, and conclusions presented in this work are entirely our own. The AI assistance was limited to improving grammar, syntax, and stylistic elements to enhance clarity and readability for an international audience.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2025.1594350/full#supplementary-material

References

Argyle, M., Lalljee, M., and Cook, M. (1968). The effects of visibility on interaction in a dyad. *Hum. Relat.* 21, 3–17. doi:10.1177/001872676802100101

Argyle, M., Lefebvre, L., and Cook, M. (1974). The meaning of five patterns of gaze. *Eur. J. Soc. Psychol.* 4, 125–136. doi:10.1002/ejsp.2420040202

Bailenson, J. N., and Blascovich, J. (2004). "Avatars," in *Encyclopedia of human-computer interaction*. (Great Barrington, MA: Berkshire Publishing Group).

Bartl, A., Wenninger, S., Wolf, E., Botsch, M., and Latoschik, M. E. (2021). Affordable but not cheap: a case study of the effects of two 3D-reconstruction methods of virtual humans. *Front. Virtual Real.* 2. doi:10.3389/frvir.2021.694617

Bartolini, T., Kresge, J., McLennan, M., Windham, B., Buhr, T. A., and Pryor, B. (1988). Perceptions of personal characteristics of men and women under three conditions of eyewear. *Percept. Mot. Ski.* 67, 779–782. doi:10.2466/pms.1988.67.3.779

Bente, G., Rüggenberg, S., Krämer, N. C., and Eschenburg, F. (2008). Avatar-mediated networking: increasing social presence and interpersonal trust in net-based collaborations. *Hum. Commun. Res.* 34, 287–318. doi:10.1111/j.1468-2958.2008.00322.x

Billinghurst, M. B., Daniel, G., Arnab, and Kiyokawa, K. (2003). Communication behaviors in colocated collaborative AR interfaces. *Int. J. Human–Computer Interact.* 16, 395–423. doi:10.1207/s15327590ijhc1603_2

Biocca, F., and Harms, C. (2002). "Networked minds social presence inventory," in *MIND labs*. Michigan: Michigan State University.

Biocca, F., and Harms, C. (2003). *Guide to the networked minds social presence inventory, 1.2.* Available online at: https://web-archive.southampton.ac.uk/cogprints.org/6743/.

Biocca, F., Harms, C., and Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence Teleoperators Virtual Environ.* 12, 456–480. doi:10.1162/105474603322761270

Borland, D., Peck, T., and Slater, M. (2013). An evaluation of self-avatar eye movement for virtual embodiment. *IEEE Trans. Vis. Comput. Graph.* 19, 591–596. doi:10.1109/TVCG.2013.24

Boucher, J. D., and Ekman, P. (1975). Facial areas and emotional information. *J. Commun.* 25, 21–29. doi:10.1111/j.1460-2466.1975.tb00577.x

Bozgeyikli, E., Bozgeyikli, L. L., and Gomes, V. (2024). "Googly eyes: exploring effects of displaying user's eyes outward on a virtual reality head-mounted display on user experience," in *2024 IEEE conference virtual reality and 3D user interfaces (VR)* (Orlando, FL: IEEE), 979–989. doi:10.1109/VR58804.2024.00117

Bozgeyikli, E., and Gomes, V. (2022). "Googly eyes: displaying user's eyes on a head-mounted display for improved nonverbal communication," in *Extended abstracts of the annual symposium on computer-human interaction in play* (Bremen Germany: ACM), 253–260. doi:10.1145/3505270.3558348

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi:10.1191/1478088706qp063oa

Braun, V., and Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qual. Res. Sport, Exerc. Health* 11, 589–597. doi:10.1080/2159676X.2019.1628806

Brugman, H., Russel, A., and Nijmegen, X. (2004). *Annotating multi-media/multi-modal resources with ELAN*. Lisbon: LREC, 2065–2068.

Burgos-Artizzu, X. P., Fleureau, J., Dumas, O., Tapie, T., LeClerc, F., and Mollet, N. (2015). "Real-time expression-sensitive HMD face reconstruction," in *In SIGGRAPH Asia 2015 Technical Briefs, SA '15* (New York, NY Association for Computing Machinery), 1–4. doi:10.1145/2820903.2820910

Canales, R., Jain, E., and Jörg, S. (2023). "Real-time conversational gaze synthesis for avatars," in *ACM SIGGRAPH conference on motion, interaction and games* (Rennes France: ACM), 1–7. doi:10.1145/3623264.3624446

Cao, C., Simon, T., Kim, J. K., Schwartz, G., Zollhoefer, M., Saito, S.-S., et al. (2022). Authentic volumetric avatars from a phone scan. *ACM Trans. Graph.* 41 (163), 1–19. doi:10.1145/3528223.3530143

Casiez, G., Roussel, N., and Vogel, D. (2012). "1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems," in *Proceedings of the SIGCHI conference on human factors in computing systems* (Austin Texas: ACM), 2527–2530. doi:10.1145/2207676.2208639

Chan, L., and Minamizawa, K. (2017). "FrontFace: facilitating communication between HMD users and outsiders using front-facing-screen HMDs," in *Proceedings of the 19th international conference on human-computer interaction with mobile devices and services* (Vienna Austria: ACM), 1–5. doi:10.1145/3098279.3098548

Chaurasia, G., Nieuwoudt, A., Ichim, A.-E., Szeliski, R., and Sorkine-Hornung, A. (2020). Passthrough+: real-time stereoscopic view synthesis for mobile mixed reality. *Proc. ACM Comput. Graph. Interact. Tech.* 3, 1–17. doi:10.1145/3384540

Chung, J. W., Fu, X. J., Deocadiz-Smith, Z., Jung, M. F., and Huang, J. (2023). "Negotiating dyadic interactions through the lens of augmented reality glasses," in *Proceedings of the 2023 ACM designing interactive systems conference* (Pittsburgh PA: ACM), 493–508. doi:10.1145/3563657.3595967

Combe, T., Fribourg, R., Detto, L., and Normand, J.-M. (2024). Exploring the influence of virtual avatar heads in mixed reality on social presence, performance and user experience in collaborative tasks. *IEEE Trans. Vis. Comput. Graph.* 30, 2206–2216. doi:10.1109/TVCG.2024.3372051

Dobre, G. C., Wilczkowiak, M., Gillies, M., Pan, X., and Rintel, S. (2022). "Nice is different than good: longitudinal communicative effects of realistic and cartoon avatars in real mixed reality work meetings," in *CHI conference on human factors in computing systems extended abstracts* (New Orleans LA: ACM), 1–7. doi:10.1145/3491101.3519628

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *J. Personality Soc. Psychol.* 23, 283–292. doi:10.1037/h0033031

Ekman, P., and Friesen, W. V. (1978). Facial action coding system. 1. doi:10.1037/t27734-000

Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., et al. (2019). Revisiting collaboration through mixed reality: the evolution of groupware. *Int. J. Human-Computer Stud.* 131, 81–98. doi:10.1016/j.ijhcs.2019.05.011

Feuchtner, T., and Müller, J. (2017). "Extending the body for interaction with reality," in *Proceedings of the 2017 CHI conference on human factors in computing systems* (Denver Colorado: ACM), 5145–5157. doi:10.1145/3025453.3025689

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol. Bull.* 133, 694–724. doi:10.1037/0033-2909.133.4.694

Friston, S. J., Congdon, B. J., Swapp, D., Izzouzi, L., Brandstätter, K., Archer, D., et al. (2021). "Ubiq: a system to build flexible social virtual reality experiences," in *Proceedings of the 27th ACM symposium on virtual reality software and technology,VRST '21* (New York, NY: Association for Computing Machinery), 1–11. doi:10.1145/3489849.3489871

Frueh, C., Sud, A., and Kwatra, V. (2017). "Headset removal for virtual and mixed reality," in *ACM SIGGRAPH 2017 talks* (Los Angeles California: ACM), 1–2. doi:10.1145/3084363.3085083

Graham, D. L., and Ritchie, K. L. (2019). Making a spectacle of yourself: the effect of glasses and sunglasses on face perception. *Perception* 48, 461–470. doi:10.1177/0301006619844680

Gugenheimer, J., Mai, C., McGill, M., Williamson, J., Steinicke, F., and Perlin, K. (2019). "Challenges using head-mounted displays in shared and social spaces," in *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (Glasgow Scotland UK: ACM), 1–8. doi:10.1145/3290607.3299028

Hauber, J., Regenbrecht, H., Billinghurst, M., and Cockburn, A. (2006). "Spatiality in videoconferencing: trade-offs between efficiency and social presence," in *Proceedings of the 2006 20th anniversary Conference on computer supported cooperative work* (Banff Alberta Canada: ACM), 413–422. doi:10.1145/1180875.1180937

He, Z., Du, R., and Perlin, K. (2020). "CollaboVR: a reconfigurable framework for creative collaboration in virtual reality," in *2020 IEEE international symposium on mixed and augmented reality (ISMAR)* Porto de Galinhas, Brazil (IEEE), 542–554. doi:10.1109/ISMAR50242.2020.00082

Ho, C.-C., and MacDorman, K. F. (2017). Measuring the Uncanny Valley effect. *Int. J. Soc. Robotics* 9, 129–139. doi:10.1007/s12369-016-0380-9

Hömke, P., Holler, J., and Levinson, S. C. (2018). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS One* 13, e0208030. doi:10.1371/journal.pone.0208030

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical J.* 50, 346–363. doi:10.1002/bimj.200810425

Ishii, H., Kobayashi, M., and Grudin, J. (1993). Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Trans. Inf. Syst.* 11, 349–375. doi:10.1145/159764.159762

Ive, J., Hoenig, J., Jaede, J., Kim, S. W., Wilson, C., III, W. A. S., et al. (2024). Wearable device for facilitating enhanced interaction

Jing, A., May, K., Lee, G., and Billinghurst, M. (2021). Eye see what you see: exploring how bi-directional augmented reality gaze visualisation influences co-located symmetric collaboration. *Front. Virtual Real.* 2, 697367. doi:10.3389/frvir.2021.697367

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., and Vogeley, K. (2018). The "social gaze space": a taxonomy for gaze-based communication in triadic interactions. *Front. Psychol.* 9, 226. doi:10.3389/fpsyg.2018.00226

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A* 32, 922–923. doi:10.1107/S0567739476001873

Kari, M., Schütte, R., and Sodhi, R. (2023). "Scene responsiveness for visuotactile illusions in mixed reality," in *Proceedings of the 36th annual ACM Symposium on user interface Software and technology* (San Francisco CA: ACM), 1–15. doi:10.1145/3586183.3606825

Kern, F., Kullmann, P., Ganal, E., Korwisi, K., Stingl, R., Niebling, F., et al. (2021a). Off-the-shelf stylus: using XR devices for handwriting and sketching on physically aligned virtual surfaces. *Front. Virtual Real.* 2, 684498. doi:10.3389/frvir.2021.684498

Kern, F., Popp, M., Kullmann, P., Ganal, E., and Latoschik, M. E. (2021b). "3D printing an accessory dock for XR controllers and its exemplary use as XR stylus," in *Proceedings of the 27th ACM symposium on virtual reality software and technology VRST '21* (New York, NY: Association for Computing Machinery), 1–3. doi:10.1145/3489849.3489949

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychol. Bull.* 100, 78–100. doi:10.1037/0033-2909.100.1.78

Kröger, J. L., Lutz, O. H.-M., and Müller, F. (2020). "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: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 international summer school, windisch, Switzerland, august 19–23, 2019, revised selected papers*. Editors M. Friedewald, M. Önen, E. Lievens, S. Krenn, and S. Fricker (Cham: Springer International Publishing), 226–241. doi:10.1007/978-3-030-42504-3_15

Kullmann, P., Menzel, T., Botsch, M., and Latoschik, M. E. (2023). "An evaluation of other-avatar facial animation methods for social VR," in *Extended abstracts of the 2023 CHI conference on human factors in computing systems* (Hamburg Germany: ACM), 1–7. doi:10.1145/3544549.3585617

Kullmann, P., Schell, T., Menzel, T., Botsch, M., and Latoschik, M. E. (2025). Coverage of facial expressions and its effects on avatar embodiment, self-identification, and uncanniness. *IEEE Trans. Vis. Comput. Graph.* 31, 3613–3622. doi:10.1109/TVCG.2025.3549887

Ladwig, P., Ebertowski, R., Pech, A., Dörner, R., and Geiger, C. (2024). Towards a Pipeline for Real-Time Visualization of Faces for VR-based Telepresence and Live Broadcasting Utilizing Neural Rendering. *J. Virtual RealityBroadcast*. 18. doi:10.48663/1860-2037/18.2024.1

Latoschik, M. E., and Wienrich, C. (2022). Congruence and plausibility, not presence: pivotal conditions for XR experiences and effects, a novel approach. *Front. Virtual Real.* 3, 694433. doi:10.3389/frvir.2022.694433

Mai, C., Knittel, A., and Hußmann, H. (2019). Frontal screens on head-mounted displays to increase awareness of the HMD users. *State Mix. Presence Collab.*

Mai, C., Rambold, L., and Khamis, M. (2017). "TransparentHMD: revealing the HMD user's face to bystanders," in *Proceedings of the 16th international Conference on Mobile and ubiquitous multimedia* (Stuttgart Germany: ACM), 515–520. doi:10.1145/3152832.3157813

Matsuda, N., Wheelwright, B., Hegland, J., and Lanman, D. (2021a). "Reverse pass-through VR," in *Special interest Group on computer Graphics and interactive techniques*

conference emerging technologies (Virtual Event: ACM), 1–4. doi:10.1145/3450550.3465338

Matsuda, N., Wheelwright, B., Hegland, J., and Lanman, D. (2021b). VR social copresence with light field displays. *ACM Trans. Graph.* 40 (244), 1–13. doi:10.1145/3478513.3480481

Mauersberger, H., Kastendieck, T., and Hess, U. (2022). I looked at you, you looked at me, I smiled at you, you smiled at me—the impact of eye contact on emotional mimicry. *Front. Psychol.* 13, 970954. doi:10.3389/fpsyg.2022.970954

McGill, M., Gugenheimer, J., and Freeman, E. (2020). "A quest for Co-located mixed reality: aligning and assessing SLAM tracking for same-space multi-user experiences," in *26th ACM Symposium on virtual reality Software and technology* (Virtual Event Canada: ACM), 1–10. doi:10.1145/3385956.3418968

Menzel, T., Botsch, M., and Latoschik, M. E. (2022). "Automated blendshape personalization for faithful face animations using commodity smartphones," in *Proceedings of the 28th ACM Symposium on virtual reality software and technology* (Tsukuba Japan: ACM), 1–9. doi:10.1145/3562939.3565622

Menzel, T., Wolf, E., Wenninger, S., Spinczyk, N., Holderrieth, L., Wienrich, C., et al. (2025). Avatars for the masses: smartphone-based reconstruction of humans for virtual reality. *Front. Virtual Real.* 6. doi:10.3389/frvir.2025.1583474

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The Uncanny Valley [from the field]. *IEEE Robotics and Automation Mag.* 19, 98–100. doi:10.1109/MRA.2012.2192811

Mori, S., Ikeda, S., and Saito, H. (2017). A survey of diminished reality: techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Trans. Comput. Vis. Appl.* 9, 17. doi:10.1186/s41074-017-0028-1

Navarro, D. (2015). *Learning statistics with R: a tutorial for psychology students and other beginners. (Version 0.6)*. Sydney, Australia.

Nowak, K. L., Fox, J., and The Ohio State University (2018). Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations on communication. *Rev. Commun. Res.* 6, 30–53. doi:10.12840/issn.2255-4165.2018.06.01.015

Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A systematic review of social presence: definition, antecedents, and implications. *Front. Robotics AI* 5, 114. doi:10.3389/frobt.2018.00114

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

Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., et al. (2016). "Holoportation: virtual 3D teleportation in real-time," in *Proceedings of the 29th annual Symposium on user interface Software and technology* (Tokyo Japan: ACM), 741–754. doi:10.1145/2984511.2984517

Pinheiro, J., Bates, D., and R Core Team (2024). *Nlme: linear and nonlinear mixed effects models*. Available online at: https://cran.r-project.org/web/packages/nlme/citation.html.

Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-effects models in S and s-PLUS*. New York: Springer. doi:10.1007/b98882

Piumsomboon, T., Dey, A., Ens, B., Lee, G., and Billinghurst, M. (2019). The effects of sharing awareness cues in collaborative mixed reality. *Front. Robotics AI* 6, 5. doi:10.3389/frobt.2019.00005

R Core Team (2024). *R: a language and environment for statistical computing*. Vienna, Austria.

Schwabe, C., and Vogt, C. (2014). *Doppelpack: mein Hund und ich*. München: Herbig.

Seele, S., Misztal, S., Buhler, H., Herpers, R., and Schild, J. (2017). "Here's looking at you Anyway!: how important is realistic gaze behavior in Co-located social virtual reality Games?," in *Proceedings of the Annual symposium on computer-human interaction in play* (Amsterdam Netherlands: ACM), 531–540. doi:10.1145/3116595.3116619

Straus, S., and McGrath, J. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. *J. Appl. Psychol.* 79, 87–97. doi:10.1037//0021-9010.79.1.87

Surkamp, C. (2014). Non-verbal communication: why we need it in foreign language teaching and how we can foster it with drama activities. *Scenario A J. Performative Teach. Learn. Res.*, 8 (2), 28–43. doi:10.33178/scenario.8.2.3

Takemura, M., Kitahara, I., and Ohta, Y. (2006). "Photometric inconsistency on a mixed-reality face," in *2006 IEEE/ACM international symposium on mixed and augmented reality* (Santa Barbara, CA: IEEE), 129–138. doi:10.1109/ISMAR.006.297804

Takemura, M., and Ohta, Y. (2002). "Diminishing head-mounted display for shared mixed reality," in *Proceedings. International symposium on mixed and augmented reality*, Darmstadt, Germany. (IEEE) 149–156. doi:10.1109/ISMAR.2002.1115084

Takemura, M., and Ohta, Y. (2005). Generating high-definition facial video for shared mixed reality," in Tsukuba International Congress Center, Tsukuba, Japan. Available online at: https://www.mva-org.jp/mva2005/.

Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., and Nießner, M. (2018). FaceVR: real-time gaze-aware facial reenactment in virtual reality. *ACM Trans. Graph.* 37, 1–15. doi:10.1145/3182644

Viola, M. (2024). Seeing through the shades of situated affectivity. Sunglasses as a socio-affective artifact. *Philos. Psychol.* 37, 2048–2072. doi:10.1080/09515089.2022.2118574

Wei, S., Bloemers, D., and Rovira, A. (2023). "A preliminary study of the eye tracker in the Meta quest Pro," in *Proceedings of the 2023 ACM international conference on interactive media experiences* (Nantes France: ACM), 216–221. doi:10.1145/3573381.3596467

Wei, S.-E., Saragih, J., Simon, T., Harley, A. W., Lombardi, S., Perdoch, M., et al. (2019). VR facial animation via multiview image translation. *ACM Trans. Graph.* 38 (67), 1–16. doi:10.1145/3306346.3323030

Wolf, D., Gugenheimer, J., Combosch, M., and Rukzio, E. (2020). "Understanding the heisenberg effect of spatial interaction: a selection induced error for spatially tracked input devices," in *Proceedings of the 2020 CHI Conference on human Factors in computing systems* (Honolulu HI: ACM), 1–10. doi:10.1145/3313831.3376876

Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., and Ishiguro, H. (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Front. Psychol.* 6, 883. doi:10.3389/fpsyg.2015.00883