Enhancing Social Experiences in Immersive Virtual Reality with Artificial Facial Mimicry

Alessandro Visconti, Davide Calandra, Federica Giorgione, and Fabrizio Lamberti

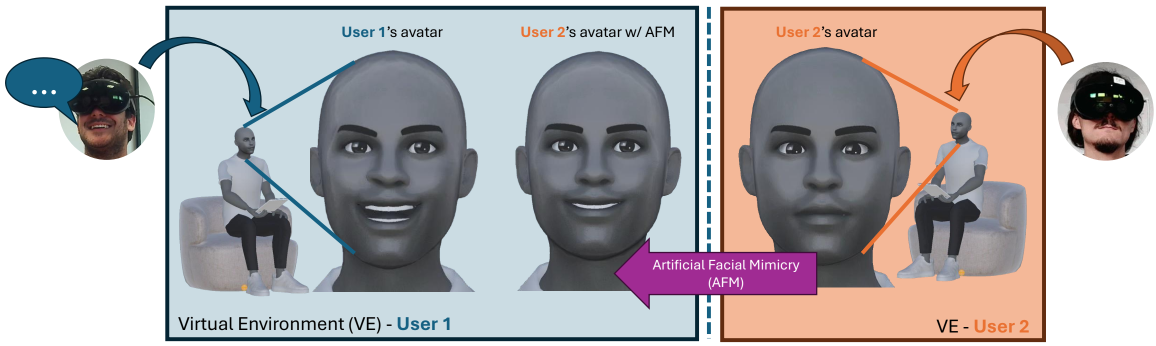

Fig. 1: Two users conversing in an immersive SVR application using the devised Artificial Facial Mimicry (AFM) system: facial expressions on the face of the right user's avatar (listening) and other non-verbal behaviors are automatically adjusted by the system based on emotions expressed by the left user (now speaking) with the aim to enhance sociality.

Abstract—The growing availability of affordable Virtual Reality (VR) hardware and the increasing interest in the Metaverse are driving the expansion of Social VR (SVR) platforms. These platforms allow users to embody avatars in immersive social virtual environments, enabling real-time interactions using consumer devices. Beyond merely replicating real-life social dynamics, SVR platforms offer opportunities to surpass real-world constraints by augmenting these interactions. One example of such augmentation is Artificial Facial Mimicry (AFM), which holds significant potential to enhance social experiences. Mimicry, the unconscious imitation of verbal and non-verbal behaviors, has been shown to positively affect human-agent interactions, yet its role in avatar-mediated human-to-human communication remains under-explored. AFM presents various possibilities, such as amplifying emotional expressions, or substituting one emotion for another to better align with the context. Furthermore, AFM can address the limitations of current facial tracking technologies in fully capturing users' emotions. To investigate the potential benefits of AFM in SVR, an automated AM system was developed. This system provides AFM, along with other kinds of head mimicry (nodding and eye contact), and it is compatible with consumer VR devices equipped with facial tracking. This system was deployed within a test-bench immersive SVR application. A between-dyads user study was conducted to assess the potential benefits of AFM for interpersonal communication while maintaining avatar behavioral naturalness, comparing the experiences of pairs of participants communicating with AFM enabled against a baseline condition. Subjective measures revealed that AFM improved interpersonal closeness, aspects of social attraction, interpersonal trust, social presence, and naturalness compared to the baseline condition. These findings demonstrate AFM's positive impact on key aspects of social interaction and highlight its potential applications across various SVR domains.

Index Terms—Immersive Environments, Virtual Reality, Artificial Facial Mimicry, Social Augmentation, Social Experiences

1 Introduction

The growing availability of consumer-grade Virtual Reality (VR) hardware and the rising interest in the Metaverse are accelerating the development of Social VR (SVR) platforms [1, 2]. These platforms enable real-time remote communication through shared Virtual Environments (VEs) [3], where users interact via digital avatars with varying levels of control depending on their VR devices. SVR platforms can transcend real-world interaction limitations through artificial augmentations, which can address sensory input gaps, enhance interactions, and support individuals with social communication challenges [4, 5]. Artificial Mimicry (AM) is one such augmentation, replicating key elements of mimicry, i.e., the unconscious imitation of verbal and non-verbal cues, which strengthens social bonds and improves communication [6–9]. AM in SVR spans behaviors such as body movements, eye contact, nodding, and facial expressions [5, 10–12].

When applied to facial expressions, AM is known as Artificial Facial Mimicry (AFM). Like in real-life interactions [13], AFM makes digital avatars adjust their facial expressions based on the emotions of others. AFM has shown potential in human-agent interactions, enhancing social presence and user engagement [14–16]. However, its use in avatar-mediated human-to-human interactions remains underexplored [5].

The simplest AFM implementation, i.e., *Emotion Mirroring*, which mirrors expressions regardless of the emotional context, can be per-

• *Alessandro Visconti, Davide Calandra, and Fabrizio Lamberti are with the Department of Control and Computer Engineering, Politecnico di Torino. E-mail: {name.surname}@polito.it*
• *Federica Giorgione is with the Politecnico di Torino E-mail: s280128@studenti.polito.it.*

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

ceived as forced or disrupt speech synchronization (Fig. 2a). Advanced AFM methods based on emotion detection can adapt mimicry dynamically, decoupling avatar movements from mimicry application. These methods include *Emotion Augmentation* to amplify the intensity of the mimicked emotion (Fig. 2b), *Emotion Camouflage* to align conflicting emotions (Fig. 2c), and *Emotion Translation* to remap expressions into peer-comprehensible forms (Fig. 2d) [10, 17].

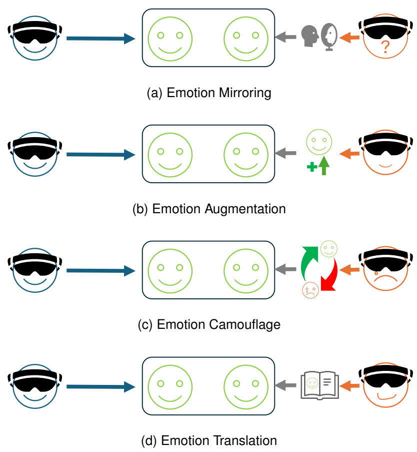

Fig. 2: Methods for applying Artificial Facial Mimicry (AFM) in immersive Social VR (SVR): User 1 (blue, left) observes User 2 (orange, right) achieving the same emotional level under different conditions: a) projecting own expression onto the peer, b) enhancing the peer's lower emotion intensity, c) replacing the peer's opposing emotion, and d) remapping expressions for better comprehension.

Although AFM has promising applications, its use to date has been largely confined to non-immersive VR, where avatar control relies on indirect methods like mouse and keyboard [12, 18]. The rise of immersive VR and SVR, supported by modern headsets with facial tracking capabilities [19–21], makes AFM more viable for direct and natural avatar control. Challenges remain, though, in replicating nuanced facial expressions, as current tracking systems often fail to capture the full range of emotions, particularly in the upper face [22]. AFM can address these gaps by enhancing expressiveness when tracking falls short.

This work explores the potential of AFM in immersive SVR to improve social interactions by influencing avatar-mediated interpersonal communication. A machine learning-based AM system was developed to manage AFM, along with mimicry of nodding and eye contact. The system dynamically adjusts the peer's avatar facial expressions by combining facial tracking data with the current user's emotional state, detected through a classifier able to identify happiness and sadness emotions, along with neutrality. When discrepancies arise between emotions, the system camouflages expressions to create matching emotions of the same intensity. During natural mimicry, if the mimicked emotion is detected at a lower intensity on the peer, the system augments it to align with the current user's intensity; otherwise, it allows the mimicry to proceed without modifying the tracking data.

2 State of the Art on Mimicry in Real-world and Virtual Interactions

Mimicry has been widely studied in real-life interactions, human-agent communication, and avatar-mediated exchanges in VEs. This section reviews mimicry literature, covering its role in real-life social dynamics, its use by virtual agents to enhance engagement, and its application in avatar-mediated interactions, highlighting gaps in SVR.

2.1 Mimicry in Real-world Interactions

Decades of research on real-life mimicry have demonstrated its role in fostering empathy, building rapport, and establishing trust between individuals [23–27]. Chartrand and Bargh [23] introduced the concept of the "chameleon effect", showing that being mimicked leads to smoother interactions, increased social affiliation, and likability. Maddux et al. [28] found that mimicry enhances trust and negotiation outcomes, while Gueguen [29] showed it promotes disclosure and honesty. In this body of work, mimicry is often studied by manipulating its levels using photos, videos, or live interactions.

The confederate paradigm, where trained individuals mirror participants, has been instrumental in studying its effects [30, 31], though it struggles to consistently link mimicry with liking, despite other positive social outcomes [32, 33]. Dyadic interactions enable natural affiliation but limit experimental control [34].

Building on these studies, mimicry emerges as a powerful mechanism to enhance persuasion, trust, and affiliation in social interactions, with promising applications in shared VEs to boost user engagement and communication [35].

2.2 Mimicry in Human-Agent Interactions

Mimicry in human-agent interactions has emerged as a prominent strategy for enhancing social engagement between humans and avatar-controlled agents [15, 16, 18, 36].

A notable example is provided by Bailenson and Yee [35], who demonstrated the persuasive power of mimicry in immersive VR by showing that virtual agents mimicking participants' head movements with a 4-second delay were perceived as more persuasive and likable, even when participants were unaware of the mimicry. This effect persisted despite the interaction being limited to a prerecorded message without verbal turn-taking.

Similarly, Verberne et al. [37] explored the impact of head movement mimicry on interpersonal trust, likability, and self-other overlap in human-agent interactions. In a route planner game, mimicry enhanced calculative trust, linked to perceived competence, but had no effect in an investment game, where relational trust, based on perceived intentions, was more critical. The rigid, delayed mimicry may have reduced agents' perceived intelligence, suggesting that more natural algorithms as well as the exploration of facial expressions are needed.

Hoegen et al. [38], extended this line of research by examining the role of facial mimicry and counter-mimicry in social interactions. Their findings revealed that agents mimicking participants' smiles and frowns promoted greater rapport and cooperation, whereas counter-mimicry showed no measurable effects. Mimicry encouraged participants to smile more, highlighting the importance of emotionally aligned reactions in human-agent dynamics.

Doo Sung et al. [39], in turn, examined the effects of non-verbal compassion, including eye contact, facial mimicry, and head-nodding, on the effectiveness of virtual counseling in a non-immersive VR setting. Participants assigned to a compassionate virtual counselor using these cues reported reduced anger and increased perceived empathy compared to those interacting with a neutral counselor. The study highlighted the potential for future implementations to incorporate technologies like face tracking and emotion recognition to enhance non-verbal interactions with virtual agents.

2.3 Mimicry in Avatar-mediated Human-to-Human Interactions

AM has been explored in avatar-mediated human-to-human interactions, aiming to replicate the benefits observed in human-agent settings. For instance, Roth et al. [10] introduced the "Mimicry Injector", a system integrating artificial non-verbal mimicry, such as body movements, into VR interactions. A negotiation study found that, while participants largely failed to detect it, AM had no significant effect on communication or negotiation outcomes, possibly due to the limited impact of body mimicry compared to facial mimicry. The authors highlighted the necessity for advanced AM techniques capable of capturing subtle non-verbal cues as well as of identifying specific triggers for AM activation (e.g., voice or gaze). Additionally, they expressed ethical concerns regarding behavioral modifications within VEs.

Early studies on AFM in human-to-human interactions were conducted in non-immersive settings. As a matter of example, Suda et al. [40] examined AFM during negotiations using avatars with partial mimicry of eye and eyelid movements. Mimicry improved cooperative outcomes, particularly in avatar-avatar interactions, but its partial implementation underscored the need for systems capable of fully replicating facial expressions while maintaining contextual appropriateness. Similarly, Suzuki et al. [41] introduced "FaceShare", a videophone system that automatically mirrors smiles using image processing. This pseudo-smile mirroring enhanced closeness and conversation quality but focused exclusively on positive expressions, highlighting the need for further exploration of negative expressions and more adaptive, context-aware mechanisms.

Nodding mimicry has also been studied. Wakabashaky et al. [12] showed that an avatar mimicking customer nodding in a non-immersive VE enhanced rapport, particularly among female participants. However, the study highlighted the influence of external factors, such as the simulated store environment, and called for the use neutral VEs to enable more precise evaluations.

Roth et al. [42] proposed the "Hybrid Avatar-Agent Technology" (HAAT), which analyzes augmentations at the dyad level by considering bidirectional adaptation and synchronization processes. The system was capable of reacting to user behavior to enhance communication and foster rapport, interpreting social signals and mapping them to the peer's reactions. Additionally, Roth et al. investigated hybrid and augmented gaze as forms of visual transformations of social phenomena using HAAT in multi-user VR environments [43] , demonstrating its capability to increase social presence [44].

Later, Roth et al. [5] advanced their earlier work by introducing "injectX", a system integrating mimicry of body movements, gaze, and facial expressions within immersive VEs. Unlike earlier systems, injectX used a decision-making process to deliver context-aware, dynamic mimicry tailored to ongoing interactions. Although a preliminary study focused on body movement blending techniques, the authors identified facial mimicry as a future research direction.

2.4 Considerations

Based on the above review, it is evident that mimicry, particularly facial mimicry, has the potential to enhance social interactions in shared VEs. While its benefits in virtual agent interactions are well-documented, research on AM and facial expressions in human-to-human interactions, particularly in immersive settings, remains limited. Previous works, such as [5], have explored AM applied to facial expressions in immersive VR but often employed simplified mimicry or lacked robust experimental validation. This paper addresses such gaps by introducing an automated AM system supporting the mimicry of nodding, eye contact, and facial expressions designed for use in immersive SVR scenarios and by evaluating its impact on human-to-human interactions.

In particular, the following hypotheses are formulated and validated through a user study:

- **H1**: Experiencing social interaction with an avatar featuring AFM enhances the quality of avatar-mediated interpersonal communication in immersive SVR.
- **H2**: Using AFM in immersive SVR does not negatively affect the perception of the behavioral naturalness of avatars.

To test these hypotheses, along with the AM system mentioned before, an immersive SVR application was developed as a test-bench, enabling two users to converse through their avatars. A study with 40 participants (20 dyads) compared an experimental group (AFM) against a control group (natural mimicry through facial tracking), with both groups experiencing AM of nodding and eye contact. Results revealed that AFM significantly enhanced avatar-mediated interpersonal communication, improving closeness, social attraction, social presence, and trust while maintaining avatar behavioral naturalness and clarity, thus positioning AFM as a valuable tool for social augmentation in SVR.

3 AM System and SVR Application

This section presents the architecture of the proposed AM system and the immersive SVR application developed as a test-bench.

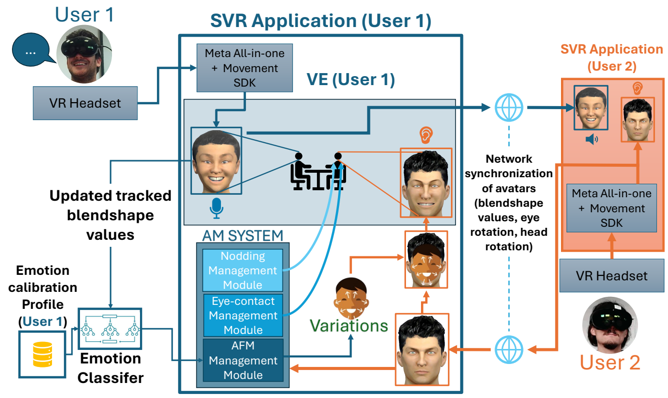

Fig. 3: Architecture of the devised AM system, illustrated by assuming that User 1 is speaking. When speech input is detected on his or her microphone, AFM is activated, and he or she can see its effects on User 2's avatar. The same applies to User 2; however, when User 2 is listening, AFM is disabled on his or her end.

3.1 AM System

The proposed AM system builds on prior research [5, 10] and can be seamlessly integrated into Unity-based multi-user VR applications. Using VR headsets with facial tracking capabilities, it captures expressions and adapts mimicry dynamically to context.

The system's architecture, illustrated in Fig. 3, consists of four main components: an *Emotion Classifier*, an *AFM Management* module, a *Nodding Management* module, and an *Eye Contact Management* module, which are detailed in the following sections. A video demonstrating its functionality is available for download[1].

3.1.1 Emotion Classifier

Instead of directly mirroring facial expressions, the AM system analyzes facial data to interpret emotions, which then guiding the AFM application of AFM.

Emotion classification uses avatars' facial blendshapes, typically employed for animation and controlled in real time via modern VR headset facial tracking SDKs. Various machine learning methods were evaluated, focusing on models implemented in the open-source *scikit-learn* library[2] that were well-suited for multi-class classification tasks. The considered models included Random Forest, Gradient Boosting, Support Vector Machine, and Logistic Regression. These models were initially trained on a trial dataset of facial blendshapes. A stratified 5-fold cross-validation was employed, followed by a grid search to compare performance. Among the models, Random Forest consistently demonstrated the best results, with accuracy ranging from approximately 78% to 98%. To limit computational cost, the Random Forest model was configured with a limited depth (10) and a low number of estimators (50). Despite these constraints, it still achieved high accuracy, with all results exceeding 90%. This configuration was chosen to maintain efficient classification on systems running simultaneously a VR application, minimizing any impact on frame rate.

The Emotion Classifier requires a preliminary calibration process, completed in a dedicated SVR application, before the AM system can be used. This process trains the classifier to recognize the user's unique facial expressions corresponding to specific emotions. The calibration consists of three phases: *Acclimatization*, *Emotion Stimulation*, and *Emotion Recording*, each set in a different VR scene.

In the *Acclimatization* phase, the user views his or her avatar in a mirror to promote embodiment and adjust to the VE. This phase is conducted only once and lasts two minutes [16]. Next, the *Emotion Stimulation* phase uses stimuli to evoke Ekman's six basic emotions (happiness, anger, disgust, fear, sadness, and surprise) along with a

[1] Sample video of AM system operation: https://bit.ly/AMsystem
[2] Scikit-learn library: https://scikit-learn.org/stable/

neutral state [45]. Drawing from Somarathna et al. [46], the application employs a mix of video clips, colors, and musical elements to elicit emotions effectively [47]. It supports media playback and customizable VE colors [48] to align with targeted emotions. Stimuli are sourced from widely used databases, such as IAPS [49] for images and the Standardized Database of Chinese Emotional Film Clips [50] for videos. In this case, a video clip associated with a specific emotion is displayed.

The user then transitions to the *Emotion Recording* phase, where he or she is placed in a VE with colors and sounds tailored to the stimulated emotion. The user is asked to verbally reflect on the emotional impact of the clip and discuss the emotions it elicited for two minutes. During this time, the avatar's facial blendshape values are continuously recorded frame by frame using the headset's facial tracking system. After the two minutes, a "Calibration Recording" file is generated, containing the captured blendshape data. The user then returns to the *Emotion Stimulation* phase, repeating the process with a new video clip for the same or a different emotion. The collection of Calibration Recording files produced during a calibration session, which may cover one or more emotions, forms the "Emotion Calibration Profile" for that user.

Once calibration is complete, the AFM Management module can operate for that user by loading his or her Emotion Calibration Profile. This profile enables the system to recognize the user's emotions and neutrality in real time, adapting mimicry behavior accordingly.

3.1.2 AFM Management

The AFM Management module controls avatar facial expressions by applying artificial modifications based on emotional states identified by the classifier. Although the module can detect any emotion the classifier has been trained on, the current implementation is limited to recognizing happiness, sadness, and neutrality with the aim to streamline system management for the purpose of the study.

Emotion recognition is based on the concept of "reaction time" [51]. Specifically, based on previous findings [52], happiness is detected when a happy facial expression is maintained for at least 0.5 seconds. Since no precise criteria exist for expressions related to sadness, the reaction time for detecting it was set to 0.6 seconds, based on average values from prior studies [51, 53]. Similarly, the reaction time for detecting neutrality and transitions from happiness or sadness to neutrality was set at 0.6 seconds.

According to existing literature, facial mimicry is observed in both speakers and listeners, though it is more perceivable on listeners [54]. Hence, the module ties AFM to speech input in order to foster emotional connection during empathetic listening [55]. When speech is detected via the microphone, the module activates AFM for the speaker. After 0.5 seconds of silence, the module classifies the user as a listener and deactivates AFM. The transition between AFM activation and deactivation (and vice versa) based on the speaking state is handled in the same way as the transitions between the other module's operating modes, as described later in this subsection.

Emotion detection via the classifier is active only when the AFM is enabled, i.e., while the user is speaking. The classifier operates exclusively on the user's avatar blendshapes. This approach removes the need to implement emotion detection on other peers' ends and transmit detection results over the network or run emotion detection for all the peers on the user's end sharing calibration profiles. The AFM can be activated as a personal setting, similar to, e.g., colorblind mode. Its functionality relies solely on the availability of peers represented by avatars with compatible blendshapes driven by facial tracking.

Fig. 4 illustrates the module's logic with User 1 acting as speaker and User 2 as listener, showing its functioning for happiness, sadness, and neutrality. The AFM module operates in three modes: *AFM Off*, *Augmentation*, and *Camouflage*, with the latter two working as defined in Section 1 and being referred later as *AFM On*. Listener's avatar modifications are applied continuously through blendshape variations updated at each frame. Unity's linear interpolation (Lerp[3]) is employed to adjusts these variations toward specific "target values" with a 50% adjustment factor, creating an exponential smoothing effect. At 90 fps, this choice ensures natural updates within perceptual thresholds, transitioning from 0 to 100 in about 0.148 seconds. Differences between modes lie in how target values are calculated based on the context.

At startup, the module initializes in *AFM Off* mode with User 1 as neutral. When the speaker's blendshape data is processed by the classifier, it detects happiness, sadness or neutrality (Fig. 4a). If the emotion is detected for at least the reaction time, the module considers it as User 1's experienced emotion. For non-neutral emotions, the system determines whether to apply AFM by comparing User 1's and User 2's blendshapes (Fig. 4b), aligning them using Action Units (AUs) of the Facial Action Coding System (FACS) [56]. This approach enables the module to move beyond a simple one-to-one mapping of users' facial expressions, facilitating the conveyance of universally recognizable emotions while preventing the replication of undesirable movements (such as those associated with speech, as discussed in previous studies [40]). As a matter of example, if the emotion detected for User 1 is happiness, according to FACS [57] for happiness alignment User 2's blendshape values related to AUs 6 and 12 (happiness) must be > 0, while AUs 1, 4, and 15 (sadness) should approach 0. Misalignment triggers AFM activation, with the module passing to *Camouflage* mode (Fig. 4c). In case of alignment, the module analyzes the blendshape values of the AUs associated with User 1's emotion, comparing those of User 1 and User 2 (Fig. 4d). If User 2's blendshapes values are higher, then User 2 is experiencing Natural Facial Mimicry (NFM) with an emotional intensity stronger than that experienced by User 1, hence *AFM Off* mode is maintained (Fig. 4e), setting target values to 0; in this way, excessive mirroring of emotional expressions is prevented, since it could even diminish trust and rapport between users [58]. Otherwise, the system goes into *Augmentation* mode (Fig. 4f). In both *Augmentation* and *Camouflage* mode, the module calculates target values for the listener's AU blendshapes related to the speaker's detected emotion as the difference between the speaker's and listener's values. In *Camouflage* mode, target values for blendshapes unrelated to the recognized emotion are set to 0 (Fig. 4g). This mechanism enhances emotional expression when users share the same emotion at varying intensities and camouflages emotions when users experience different emotional states, promoting smoother social interactions [59]. If User 1 is detected as neutral, the module passes to *AFM Off* mode.

It should be noted that the system does not employ emotion translation as defined in Section 1. While it uses FACS AUs to interpret the peer's expressions and control AFM application, true translation would require detecting the peer's emotion via his or her Emotion Calibration Profile and mapping it to a standard emotion. This is not possible in the current implementation, as the classifier remains inactive for the listening peer, and detection results are not shared between users.

3.1.3 Nodding and Eye-Contact Management

Eye contact is a key non-verbal cue that signals awareness and enhances social connection, while nodding indicates engagement in conversations [60]. Building on these principles and on prior studies [12, 39], the AM system incorporates artificial mimicry for eye contact and nodding.

For eye contact, the system detects when a user looks at another's avatar using raycasts from the eyes that intersect with a sphere collider on the peer's avatar head. If the gaze is not reciprocated, the system adjusts the peer's avatar to simulate mutual eye contact, using interpolated eye rotations completed within 1 second. This mirroring persists for up to 9 seconds, the optimal duration for conversational eye contact [39].

For nodding, the system uses the open-source *HeadNodYesSensor.cs* Unity script [61] to detect head movements. When a nod is identified, the peer's avatar mirrors it, with head rotation angles randomly set between 15 and 20 degrees to avoid unnatural imitation.

The combination of the Eye-Contact, Nodding, and AFM Management modules forms the AM system, which can be integrated into multi-user immersive SVR applications that support facial tracking and standard avatar blendshapes.

3.2 SVR Application

The AM system was developed to study the benefits of AFM in human-to-human interactions within immersive SVR platforms. Hence, as

[3]Unity - Scripting API: Mathf.Lerp: http://tiny.cc/eb34001

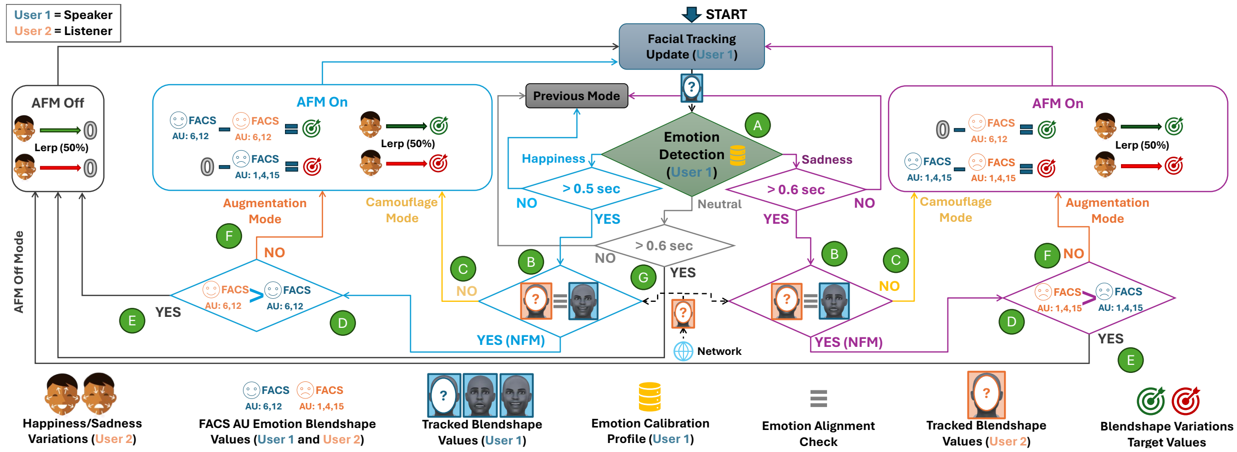

Fig. 4: Overview of the AFM Management module's operational logic, demonstrating the interaction between User 1 (speaker) and User 2 (listener) in three emotional states: happiness, sadness, and neutrality. User 1's blendshapes are directly tracked via facial tracking, while User 2's blendshapes are synchronized over the network. The letters A to G are those used in Section 3.1.2.

said, for the purpose of running the experiments an immersive VR application was created as a test-bench whereto deploy the system.

The application was developed in Unity 2021.3, utilizing the Meta XR All-in-One SDK for immersive VR functionalities, including eye and facial tracking, and the Mirror networking library[4] to support multi-user environments with synchronized avatars. Real-time voice communication was implemented using the Dissonance Voice Chat asset[5]. Special focus was placed on avatar design, VE simplicity, and seamless integration of the AM system. Full-body humanoid avatars were created using Ready Player Me[6], balancing high fidelity with low polygon count and providing facial blendshapes compatible with the Meta XR SDK. Two avatars, male and female, were designed in a similar style with black and white textures to avoid the presence of distinguishing features and minimize biases, addressing known tendencies for ingroup favoritism and outgroup discrimination in real life [62] and in SVR platforms [63, 64]. The application places two users in a virtual room, seated across from each other to focus on the face while minimizing distractions from body movements. Avatar control is restricted to head, face, and eye movements, managed via the Meta Movement SDK[7]. Avatar facial expressions are adjusted using 52 blendshapes automatically updated by the SDK based on facial tracking inputs.

3.3 Virtual Environment

In the immersive SVR application, a simplified yet realistic environment was used to create a relaxed, social atmosphere conducive to experimentation [65, 66].

The environment included minimal elements (a door, a window, and a small table) to reduce potential distractions [12]. Users were seated in armchairs across from each other, each with a virtual notebook positioned for easy reading. During the experiment, the notebook displayed conversation-guiding questions, as detailed in Section 4. To prevent visual clipping of the user's own avatar, the corresponding mesh was disabled, while the peer's avatar was shown in an idle seated animation, holding the notebook.

The application also included an acclimatization scene to help users adjust to the experience, whose environment had the same configuration used during calibration.

4 Evaluation of AFM in SVR: Experiment Overview

As previously stated, the immersive application and the automatic AM system were used to design a user study evaluating the benefits of facial mimicry in multi-user SVR contexts. The study employed a between-dyads design, dividing participant pairs into two equal groups: one using AFM and the other serving as a control group. To align with best practices and isolate the impact of facial mimicry, both the groups included eye contact and nodding as consistent non-verbal cues.

Each dyad included two participants engaging in a conversation. While using a confederate in place of one participant could have been an alternative, this configuration was avoided to preserve experimental validity, as repeated participation by a confederate might lead to inconsistent levels of sociability [32, 33]. For the experimental group, the AFM system was activated for one participant in each dyad at a time, applying it to the listener's avatar during the speaker's turn. When the roles reversed, the system switched to the new listener. To facilitate the conversation, the Fast Friends Protocol (FFP) was employed, a method widely recognized in digital sociality studies [67, 68] for its effectiveness in fostering closeness between strangers. The FFP comprises three sets of 12 progressively personal questions designed to be answered over a 45-minute session. In particular, a reduced version of the protocol was implemented to better suit the context.

The hardware setup included two Meta Quest Pro VR headsets with integrated facial and eye tracking, connected via Meta Quest Link in a tethered setup. The headsets were paired with two Intel Core i7-12700 PCs (16GB RAM, GeForce RTX 3070 Ti), connected to the same wired local network. To ensure privacy, the PCs were placed in separate rooms far enough apart to prevent users from hearing each other through the walls.

4.1 Emotion Calibration Profile Creation

To create the Emotion Calibration Profile for each participant, it was necessary to define the target emotions and the ways to evoke them. The focus was on affiliative emotions that are easily stimulated and enduring, which led to the selection of happiness (joy, fun) and sadness.

Publicly available emotional stimuli databases were analyzed to determine suitable elicitation methods. For the *Emotion Stimulation* phase, the AVDOS dataset [69] was selected. This dataset contains 60 video clips, each lasting 60 seconds, categorized as positive, negative, or neutral. The ten clips with the highest positive valence and the ten with the lowest negative valence were shortlisted, and six were ultimately selected. Each participant viewed two clips associated with happiness, two with sadness, and two with neutrality.

For the *Emotion Recording* phase, two color profiles were applied to adjust the VE tones. An happy profile increased brightness and introduced warmer colors to evoke happiness, while a sad profile reduced brightness and emphasized cooler tones to induce sadness. Audio stimuli were also incorporated, following previous research [70]. A major chord composition (Vivaldi's Allegro molto in C major) was played in a loop for happiness, while a minor chord composition (Sainte

[4]Mirror Networking for Unity: http://tiny.cc/jndmzz
[5]Dissonance Voice Chat for Unity: http://tiny.cc/cndmzz
[6]Ready Player Me: https://readyplayer.me/it
[7]Meta Movement SDK for Unity: http://tiny.cc/aynmzz

Colombe's Les pleurs) was used for sadness. Each participant's Emotion Calibration Profile was composed of six Calibration Recording files, corresponding to the emotions elicited during the process.

4.2 Participants

The study involved 40 participants (24 males, 16 females) aged 19 to 33 years ($M = 25.56$, $SD = 2.75$). The sample comprised students and staff from the authors' university who volunteered for the experiment, representing potential future users of the SVR application. On a 1-to-5 scale (1="Never used," 5="I use it every day"), participants had moderate prior VR experience ($M = 2.44$, $SD = 1.27$) but were generally unfamiliar with SVR ($M = 1.28$, $SD = 0.57$).

Participants were divided into two groups with no statistically significant differences in age or VR/SVR experience. The *AFM* group experienced the immersive SVR application with the full AM system enabled including AFM, whereas the *Control* group experienced the AM system with only nodding and eye contact. None of the participants were aware that the AM was in place during the experiment. To minimize biases, participants were paired with strangers, and each pair was grouped by gender, consisting of either two males or two females, as done in similar studies [67, 68, 71].

4.3 Procedure

The experimental procedure, consisted of the following phases: 1) Pre-experience questionnaire, 2) Calibration, 3) Acclimatization in the VE, 4) Conversation, 5) Post-experience questionnaire.

4.3.1 Pre-experience Questionnaire

Following a brief explanation of the study procedure, participants completed a pre-experience questionnaire to gather demographic information, prior experience with relevant technologies, and personality traits.

4.3.2 Eye-tracking Calibration and Loading of the Emotion Calibration Profile

Each participant was equipped with a VR headset and first underwent eye-tracking calibration for the Meta Quest Pro. The participant was then immersed in a VR scene to complete the classifier setup process (Section 3.1.1) using the stimuli described in Section 4.1. Once the setup was complete, the classifier was activated and loaded with the participant's Emotion Calibration Profile, preparing for the actual immersive SVR experience with the AM system.

Dyads were formed based on compatibility from the pre-experience questionnaire, and participants were occasionally asked to return later (the same day or another) if a suitable partner was unavailable. During the experiment, the two participants were in separate rooms, supervised by different experimenters.

4.3.3 Acclimatization in the Immersive SVR Application

Before entering the VE and beginning the experiment, participants engaged with the single-user acclimatization scene described in Section 3.3. Following similar approaches in prior studies [10, 16], they were given 2–3 minutes to experience embodiment and familiarize themselves with the avatar by observing how it portrayed their facial expressions in the virtual mirror.

4.3.4 Conversation

Following acclimatization, participants entered the actual SVR application, a multi-user scene where they conversed for 20 minutes. Conversation topics were based on a subset of FFP questions, a methodology previously used in [67, 68, 72, 73].

To maintain immersion while allowing participants to read the questions naturally, each avatar was equipped with a virtual notebook displaying the prompts. Participants were instructed to take turns speaking without interruption. One participant would begin by reading the first question and answering it. Afterwards, the other participant would answer the same question; he or she would then read and answer the next question, alternating in this manner throughout the conversation.

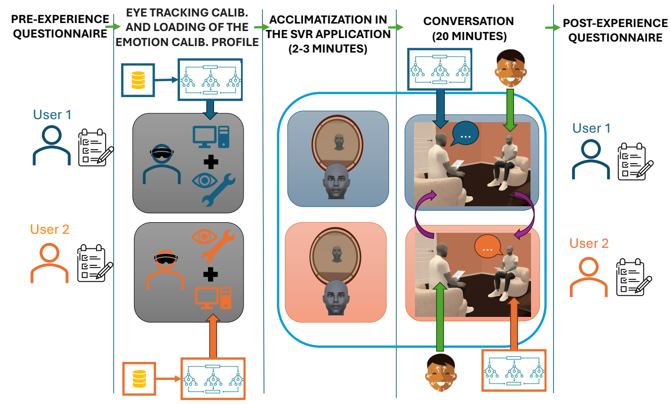

Fig. 5: Phases of the experimental procedure.

4.3.5 Post-experience Questionnaire

Lastly, after the 20 minutes conversation, the participants were administered a post-experience questionnaire.

4.4 Measures

The experimental evaluation included both subjective and objective measures. Subjective data were collected through pre- and post-experience questionnaires, while objective data were obtained from AM system logs, including detected emotions and information on the application of the AM.

The pre-experience questionnaire gathered participants' demographics, familiarity with VR, and prior SVR experience, with the latter two dimensions assessed on a 5-point Likert scale. It also included the abbreviated Big Five Inventory (BFI-S) [74], which helped to identify differences in mimicry perception based on social tendencies and was key in pairing participants with similar personality traits (e.g., introverted or extroverted) for the experimental groups.

The post-experience questionnaire included standard tools to assess various dimensions of interpersonal communication. Previous studies explored aspects such as interpersonal closeness [72], social presence [75], interpersonal attraction [76], and interpersonal trust.

To measure interpersonal closeness and connection, the Inclusion of Others in the Self (IOS) tool was employed. Widely used in FFP-based studies [72], the IOS scale asks participants to select one of seven images depicting increasingly overlapping circles, representing levels of perceived closeness on a 1-to-7 scale.

To compare the two configurations in terms of social presence, the Networked Minds Measure Scale (NMMS) [77] was used. This tool assesses co-presence, attentional involvement, perceived emotional contagion, comprehension, and behavioral interdependence, with responses on a 1-to-7 scale.

Following the methodology proposed by McCroskey and McCain [76], two dimensions of interpersonal attraction, i.e., social and task attraction, were evaluated. Social attraction was measured using four statements from Davis and Perkowitz [78], while task attraction was assessed with four statements from Burgoon [79]. Participants rated their agreement with each statement on a 7-point Likert scale (1="Strongly disagree", 7="Strongly agree").

Additional tools were used to evaluate the realism of the avatars and the behavior of the two configurations, ensuring that issues like the uncanny valley phenomenon [80] were avoided. The perception of the other participant's verbal and non-verbal cues was assessed using the behavioral naturalness section of the questionnaire by Kullmann et al. [81], which helped to spot discrepancies between expected and actual behavior in the immersive SVR application (1-to-7 scale). Additionally, the Godspeed questionnaire [40], widely used in virtual agent and robot development, was employed to evaluate the realism of the avatars' behavior (1-to-5 scale).

Finally, as frequently done in similar contexts [75], interpersonal trust was assessed using custom questions, as no standardized questionnaire exists for this purpose. Dimensions like communication satisfaction and attention to behavioral cues were excluded due to misalignment with the study protocol or redundancy with other measures. The custom questions, rated on a 1-to-7 scale, focused on trust, empathy, and sense of being emotionally understood by others.

The questionnaires, administered in English without altering the standard wording, are available for download alongside the adapted FFP questions provided in both English and Italian (the language used during the experiment)[8].

Regarding the objective measurements, data were collected to evaluate the functioning of the AM system. They included the number of nods detected, the time each participant spent looking at the peer's face, the detection of happiness, sadness, and neutrality, as well as the distribution of operating time across the various modes of the AFM Management module.

5 Results

This section presents the results obtained from the experimental activity.

For the post-experience questionnaire, the Shapiro-Wilk test was used to evaluate data normality. Since data were found to be non-normally distributed, the non-parametric Mann-Whitney test was adopted to assess statistical significance, with a threshold $p < .050$. Additionally, the reliability of the measures for each group was evaluated using Cronbach's alpha, and effect sizes were calculated using Cohen's d.

Starting with interpersonal closeness, participants in the AFM group reported significantly higher IOS scores compared to those in the Control group ($M = 4.81, SD = 0.81$ vs. $M = 4.06, SD = 0.97; d = 0.84$). These results are illustrated in Fig. 6.

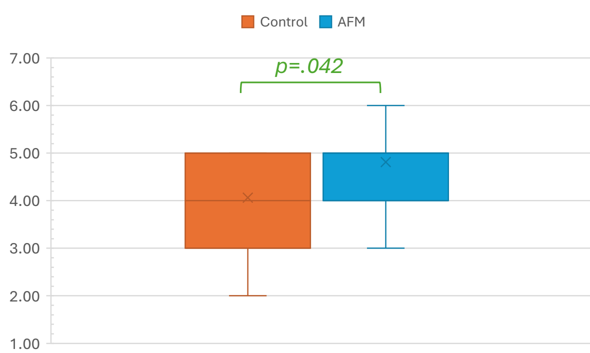

Fig. 6: Box plot showing the distribution of the scores for the Inclusion of Others in the Self (IOS) metric, assessing the interpersonal closeness. Bracket indicates a statistically significant difference.

For the Networked Minds Measure Scale (NMMS), internal consistency was acceptable, with Cronbach's $\alpha = .95$ for the AFM group and $\alpha = .91$ for the Control group. Results are presented in Fig. 7. Significant differences were observed in three out of the five sections: co-presence ($M = 6.05, SD = 0.56$ vs. $M = 5.16, SD = 0.95; d = 1.13$), perceived emotional contagion ($M = 5.09, SD = 1.00$ vs. $M = 4.13, SD = 1.06; d = 0.93$), and behavioral interdependence ($M = 4.96, SD = 0.94$ vs. $M = 3.90, SD = 1.22; d = 0.98$).

Analyzing the individual sections of the NMMS, a significant difference was found in the co-presence section for the question related to the feeling of not actually being in the same room as the other ($M = 1.94, SD = 0.97$ vs. $M = 3.19, SD = 1.59; p = .011; d = 0.95$), where the AFM group perceived the other as physically closer than the Control group. Similar results were obtained for the mirrored question about participants' perception of how much the other felt like they were in a different place, again in favor of the AFM group ($M = 1.81, SD = 0.63$ vs. $M = 3.25, SD = 1.68; p = .009; d = 1.13$). A similar pattern was observed for the two questions related to mutual awareness, with the

[8]Questionnaire and FFP questions: https://bit.ly/3BU9JnB

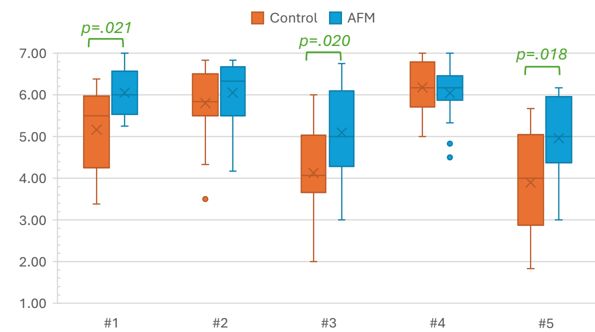

Fig. 7: Box plot showing the distribution of the scores for the Networked Minds Measure Scale (NMMS) (#1 Co-presence, #2 Perceived attentional engagement, #3 Perceived emotional contagion, #4 Perceived comprehension, #5 Perceived behavioral interdependence). Brackets indicate statistically significant differences.

AFM group reporting significantly higher scores. One question focused on how often the participant felt that the other was aware of him or her ($M = 6.06, SD = 0.83$ vs. $M = 5.12, SD = 1.05; p = .016; d = 0.99$). The other question analyzed how frequently the participant believed the other perceived to be aware of him or her ($M = 5.63, SD = 0.93$ vs. $M = 4.75, SD = 1.09; p = .040; d = 0.87$). For the perceived emotional contagion, the AFM group assigned higher ratings for both happiness contagion ($M = 5.75, SD = 0.90$ vs. $M = 4.63, SD = 1.36; p = .024; d = 0.97$) and sadness contagion ($M = 5.06, SD = 1.43$ vs. $M = 3.56, SD = 1.27; p = .008; d = 1.11$). The same trend was observed for questions related to perceived behavioral interdependence. The participants in the AFM group felt they influenced the other's actions ($M = 5.00, SD = 1.37$ vs. $M = 3.75, SD = 1.15; p = .005; d = 0.99$) significantly more than those in the Control group. They also perceived that the other's actions were dependent on their own ($M = 4.88, SD = 1.36$ vs. $M = 3.56, SD = 1.32; p = .005; d = 0.98$) and that the other's behavior was a direct response to theirs ($M = 5.06, SD = 0.97$ vs. $M = 3.88, SD = 1.41; p = .020; d = 0.98$) more than those in the Control group. No significant differences were found regarding perceived attentional engagement and comprehension.

Regarding interpersonal attraction, as shown in Fig. 8 social attraction was significantly higher for the AFM group compared to the Control group, particularly for likability ($M = 6.13, SD = 1.05$ vs. $M = 5.00, SD = 1.50; d = 0.87$) and friendliness ($M = 6.56, SD = 0.61$ vs. $M = 5.69, SD = 1.04; p = .014; d = 1.02$). The aggregate of the section shows a significant difference in favor of AFM group ($M = 6.20, SD = 0.71$ vs. $M = 5.34, SD = 1.30; p = .041; d = 0.82$). No significant differences were found for the single items of task attraction section and for the aggregate ($M = 6.18, SD = 0.90$ vs. $M = 5.76, SD = 0.90; p = .150; d = 0.47$). The internal consistency of the statements was acceptable, with Cronbach's $\alpha = .94$ for the AFM group and $\alpha = .93$ for the Control group.

Behavioral naturalness was rated in general significantly higher by the AFM group ($M = 5.64$, $SD = 0.68$ vs. $M = 5.16$, $SD = 0.66$, $p = .050$, $d = 0.71$), although no significant differences were found for the single items of the questionnaire. The internal consistency of the questionnaire was good for the AFM group, with Cronbach's $\alpha = .83$, but low for the Control group, with $\alpha = .66$. The Godspeed questionnaire revealed higher ratings for avatar realism in the AFM group ($M = 4.05$, $SD = 0.72$ vs. $M = 3.33$, $SD = 1.00$, $p = .040$, $d = 0.83$), particularly for the perception of a less artificial behavior $M = 4.13, SD = 1.11$ vs. $M = 3.00, SD = 1.17; p = .001; d = 0.99$), as shown in Fig. 9. The internal consistency of the questionnaire was acceptable, with Cronbach's $\alpha = .73$ for the AFM group and $\alpha = .85$ for the Control group.

Regarding custom questions related to interpersonal trust, significant differences in favor of the AFM group were observed for all

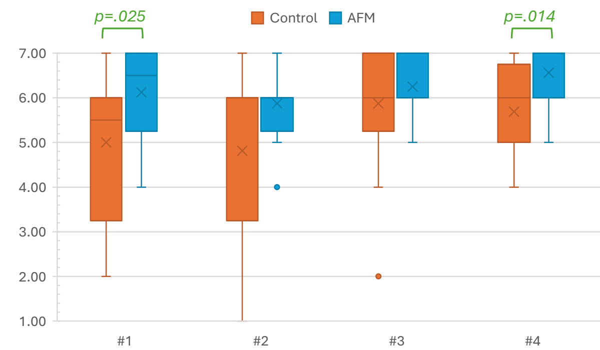

Fig. 8: Box plot showing the distribution of the scores for the statements of the questionnaire on social attraction [78] (#1 I like my partner, #2 I would get along well with my partner, #3 I would enjoy a casual conversation with my partner, #4 My partner is friendly). Brackets indicate statistically significant differences.

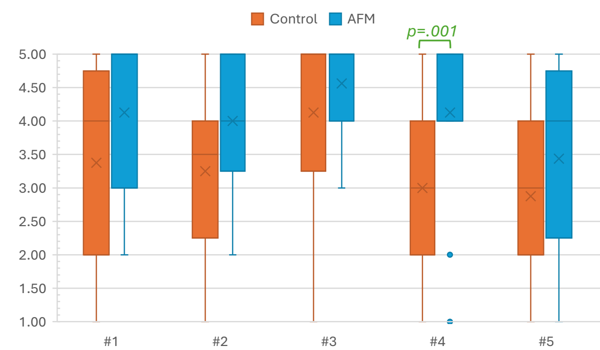

Fig. 9: Box plot showing the distribution of the scores for the items of the anthropomorphism section of the Godspeed questionnaire (#1 Fake/Natural, #2 Machinelike/HumanLike, #3 Unconscious/Conscious, #4 Artificial/Lifelike, #5 Moving rigidly/Moving elegantly). Brackets indicate statistically significant differences.

the questions, as shown in Fig. 10. In particular, the participants stated to feel more comfortable with the other ($M = 6.31, SD = 0.46$ vs. $M = 5.56, SD = 0.99; d = 0.96$), found to be better understood both when they felt happy ($M = 6.19, SD = 0.73$ vs. $M = 5.19, SD = 1.18; d = 1.02$) and when they felt sad ($M = 5.81, SD = 1.01$ vs. $M = 4.75, SD = 0.75; d = 1.19$), experienced a higher sense of trust in the other ($M = 5.69, SD = 0.92$ vs. $M = 4.88, SD = 0.86; d = 0.92$), and perceived the other as more empathetic ($M = 6.06, SD = 0.75$ vs. $M = 5.13, SD = 0.86; d = 1.17$). The internal consistency of the questions was acceptable, with Cronbach's $\alpha = .85$ for the AFM group and $\alpha = .83$ for the Control group.

To verify the absence of significant differences between pairs in each condition, a Nested Ranks Test (a Mann-Whitney-Wilcoxon variant for nested ranks[9]) was conducted due to non-normal sample distributions. The analysis revealed only a few significant differences ($p < .050$) limited to specific items, primarily within the Control group. A variability assessment was subsequently performed on these items, using the Coefficient of Variation (CV) with 20% as threshold for moderate to high variability [82]. Significant results included the perception of being in the same room ($p = .031$; CV: 33.32% vs. 20.35%), sadness emotional contagion ($p = .010$; CV: 20.35% vs. 19.58%), personal liking of the partner ($p = .038$; CV: 26.16% vs. 14.58%), and perception of the partner's appearance as artificial versus lifelike ($p = .027$; CV: 37.44% vs. 17.68%).

Objective data from the AM system logs revealed no significant differences in terms of nodding ($M = 8.58, SD = 5.36$ vs. $M = 11.00, SD = 7.77; p = .620; d = 0.36$) and eye-contact ($M = 87.00, SD = 3.22$ vs. $M = 85.20, SD = 3.49; p = .150; d = 0.54$). Moreover, Table 1 provides a detailed breakdown of AFM usage in the various operational states, categorized by speaker's detected emotion.

[9]nestedRanksTest: https://github.com/douglasgscofield/nestedRanksTest

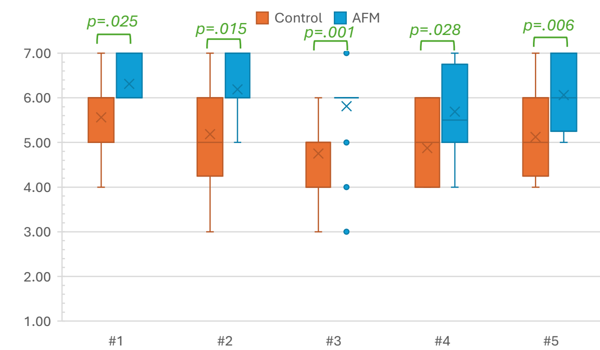

Fig. 10: Box plot showing the distribution of the scores for the custom questions assessing interpersonal cloneness (#1 The other person's reactions made me feel comfortable, #2 When I was happy I felt understood, #3 When I was sad I felt understood, #4 I think the other person deserves my trust, #5 I think the other person was empathic with me). Brackets indicate statistically significant differences.

Det. Emot.	**(a) Time Det.**	**(b) Augm.**	**(c) Camo.**	**(d) AFM On**	**(e) AFM Off (b + c)**	**(f) NFM**
Happy	32.68 (20.99)	22.96 (15.05)	1.07 (1.15)	24.04 (15.24)	8.65 (5.87)	30.48 (19.81)
Sad	55.06 (18.74)	29.34 (10.77)	3.19 (3.54)	32.53 (11.57)	22.54 (17.36)	48.96 (15.85)
Neutr.	12.25 (2.38)	–	–	–	12.25 (2.38)	–
Total	100.00 (–)	52.30 (15.26)	4.27 (3.03)	56.57 (14.59)	43.43 (14.59)	79.45 (4.71)

Table 1: Average percentages of AFM usage across all participants during the time they acted as speakers in the experiment (∼10 minutes), categorized by detected emotions (happiness, sadness, or neutral). Standard deviations are reported in parentheses. Time a given emotion has been detected is reported (a). The AFM module operates in three modes: Augmentation (b), Camouflage (c), and AFM Off (e), with their combined usage totaling 100% (b + c + e = 100%). NFM (f) represents the percentage of time the peer's emotion naturally aligned with the speaker's emotion, regardless of AFM application.

6 Discussion

The study reported in this paper investigated two hypotheses: AFM can enhance the quality of avatar-mediated interpersonal communication in immersive SVR (**H1**) and does not negatively affect the perception of avatars' behavioral naturalness (**H2**). The results provide substantial support for both the hypotheses, highlighting AFM's positive impact on social interaction and avatar realism in immersive VEs.

With respect to results supporting **H1**, participants in the AFM group reported higher levels of interpersonal closeness. This finding aligns with prior research on mimicry, which emphasized its role in fostering social bonds in real-life interactions [34], remote communication [41], and human-agent interactions [39, 83]. Moreover, this outcome is supported also by the comments shared by two participants in the Control group: one noted that the peer seemed to offer little consideration, while another expressed a lack of genuine connection during the conversation. With regard to the two aspects of interpersonal attraction, higher ratings for social attraction (in terms of likability and friendliness) with AFM further reinforce its role in enhancing participants' perception of their peer's presence and responsiveness, as observed in similar contexts of human-agents interaction [35, 37]. However, consistently with prior studies on interpersonal communication in SVR [75], no significant differences were observed for answers to questions on communication satisfaction, with both the groups achieving excellent results. Similarly, no significant differences were found for the statements related to task attraction. These outcomes could potentially be due to the engaging nature of the conversation topics and task structure. Regarding interpersonal trust, AFM contributed to stronger perceptions of peer's reliability. This result mirrors previous findings in psychology and virtual agent studies, which showed that mimicry can foster trust and cooperation [28, 37]. Similarly, AFM demonstrated its capacity to promote empathy, particularly for happiness and sadness, as observed in previous studies on mimicry in physical reality [84] and with virtual agents [39]. The AFM group also exhibited higher social presence, particularly in terms of co-presence, perceived emotional contagion, and behavioral interdependence. The significant differences in co-presence ratings suggest that AFM can enhance social interaction in SVR, leading participants to perceive their peer as more present, aware, and confident with them [85]. Emotional contagion results validate the system's effectiveness in conveying emotions on avatars, while behavioral interdependence highlights AFM's ability to amplify the perception of mutual responsiveness in actions and expressions. These findings underscore how AFM, combined with nodding and eye contact, can effectively simulate natural social dynamics in immersive environments. Concerning perceived attentional engagement and comprehension when using AFM, participants found the experience adequate, with no improvement in terms of understanding. This finding is interesting since it indicates that AFM did not cause distractions or diminish participants' focus on the peer, which could have impaired comprehension. Although verbal communication likely played a key role in attention and comprehension during the task, prior studies indicate that non-verbal cues significantly contribute to communication, accounting for the majority of interactions [71, 86].

The absence of distractions also indicates that AFM did not negatively affect the behavioral naturalness of avatars, supporting hypothesis **H2**. Participants in the AFM group consistently rated avatars as more natural and lifelike, with no reported disturbance or discomfort. These results align with prior studies on mimicry in VEs [40], suggesting that AFM can enhance realism without triggering, e.g., the uncanny valley effect [87]. AFM's ability to subtly amplify facial cues likely contributed to this perceived authenticity, fostering synchronized interactions that align with participants' expectations. Moreover, the effectiveness of the 3D avatar models, which avoided being overly realistic or artificial, further supported the positive reception of the system. These considerations on realism and naturalness are also supported by comments from participants in the AFM group, who did not perceive the presence of artificial manipulations during the experiment and expressed surprise when informed afterward that AFM had been applied to the peer.

Objective measures derived from the AM system logs revealed that AFM remained active for just over 50% of the time, suggesting that the system did not fully replace the participant's natural behaviors but rather functioned as an additional tool to enhance emotional mimicry during the interaction. During most of its active time, the system operated using emotion augmentation, amplifying the listener's mimicked emotion and increasing periods of emotional alignment between participants, potentially enhancing communication [59]. The prevalence this functioning mode correlates with an important presence of NFM, which reflects the subconscious emotional alignment observed in real-life interactions [88]. The limited use of emotion camouflage further suggests that participants naturally aligned emotionally with each other. Notably, sadness was frequently detected, likely due to the predominantly sad tone of the FFP questions used in the experiment.

The results of the nested analysis further highlight the robustness of the above findings since significant variability was primarily observed in the Control group, suggesting that AFM provided a more consistent and reliable user experience. For what it concerns same-room perception, the high variability in the both groups indicates that the difference was not due to greater consistency in the particular group. Regarding the emotional contagion of sadness, the Control group showed marginally higher variability than the AFM one. For the personal liking of the partner, the greater variability in the Control group likely reflects the subjective nature of the measure and the absence of preference-based pairing. Finally, the higher variability in the Control group for what it concerns partner appearance could be attributed to differences in facial tracking performance and familiarity with 3D avatars [89].

7 Conclusions

This paper demonstrated the potential of AFM in enhancing avatar-mediated interpersonal communication within SVR platforms. The experimental findings showed that AFM can improve closeness, aspects of social attraction, trust, and social presence, specifically in terms of co-presence, perceived emotional contagion, and perceived behavioral interdependence. Such improvements were achieved while preserving the behavioral naturalness of avatars. These results position AFM as a valuable tool for social augmentation in the considered context.

Several limitations must be acknowledged. The sample size, while sufficient for initial analysis, restricts the generalizability of the findings to broader populations. Future studies should address this point by involving larger and more diverse participant pools. Additionally, the focus on a basic set of emotions, i.e., happiness and sadness, leaves room for experimenting AFM with other emotions, such as anger, fear, surprise, and disgust. Exploring these emotions could provide a deeper understanding of the system's versatility across diverse social contexts. In this respect, a key challenge for future work will be to accurately design calibration videos that effectively evoke distinct emotions. This approach would facilitate the creation of well-defined Emotion Calibration Profiles, enabling the classifier to detect a broader range of emotions with greater accuracy. Another limitation lies in the emphasis on facial mimicry alone, excluding other non-verbal cues, such as body posture and gestures, which play an integral role in communication. Moreover, while the between-dyad design was chosen to avoid carryover effects, it might not fully capture individual variability in response to AFM, which could be addressed by employing within-dyad designs in future studies.

Building on current findings and on considerations above, future work should aim to integrate AFM with other non-verbal communication modalities, such as gestures, body language, and vocal intonation. A multimodal approach would create a more comprehensive framework for fostering social connections in immersive environments. Expanding the range of emotions considered, including culturally dependent ones, would also offer deeper insights into AFM's impact across diverse populations. Furthermore, longitudinal studies would be key to understand the long-term effects of AFM on user behavior, interaction dynamics, and emotional well-being. Such studies could also evaluate the potential benefits of AFM for specific groups, such as individuals with social communication challenges or those at risk of social isolation.

Ethical considerations must be addressed too. The subtle influence of AFM on user perception and behavior raises concerns about transparency and potential for manipulation, particularly in commercial applications [5, 10]. Regulations should mandate that users are fully informed about the system's functions and retain control over its activation. Additionally, the long-term psychological effects of AFM, such as the risks of dependency or desensitization, must be explored further. Balancing the transformative potential of AFM with safeguards to maintain authenticity and protect user autonomy is indeed crucial.

In conclusion, AFM represents a powerful tool for enhancing interpersonal dynamics in SVR. By addressing current limitations and exploring its full potential, AFM can revolutionize social interactions in VEs while adhering to ethical principles and fostering positive outcomes for diverse users.

Acknowledgments

This work has been supported by the VR@POLITO initiative.

References

[1] T. J. Tanenbaum, N. Hartoonian, and J. Bryan, "How do I make this thing smile? An inventory of expressive nonverbal communication in commercial social virtual reality platforms," in *Proc.\ CHI*, CHI '20, pp. 1–13, 2020. 1

[2] H. J. Smith and M. Neff, "Communication behavior in embodied virtual reality," in *Proc.\ CHI*, CHI '18, pp. 1–12, 2018. 1

[3] D. Maloney, G. Freeman, and D. Y. Wohn, "Talking without a voice: Understanding non-verbal communication in social virtual reality," *Proc.\ HCI*, vol. 4, pp. 1–25, 2020. 1

[4] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk, "Transformed social interaction, augmented gaze, and social influence in immersive virtual environments," *Human Communication Research*, vol. 31, no. 4, pp. 511–537, 2005. 1

[5] D. Roth, G. Bente, P. Kullmann, D. Mal, C. F. Purps, K. Vogeley, and M. E. Latoschik, "Technologies for social augmentations in user-embodied virtual reality," in *Proc.\ VRST*, VRST '19, pp. 1–12, 2019. 1, 3, 9

[6] M. Stel and R. Vonk, "Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction," *British J. of Psychology*, vol. 101, no. 2, pp. 311–323, 2010. 1

[7] T. L. Chartrand and R. van Baaren, "Chapter 5 human mimicry," in *Adv. in Experimental Social Psychology*, vol. 41, pp. 219–274, Academic Press, 2009. 1

[8] U. Hess and A. Fischer, "Emotional mimicry: Why and when we mimic emotions," *Social and Personality Psychology Compass*, vol. 8, no. 2, pp. 45–57, 2014. 1

[9] K. L. Poole and H. A. Henderson, "Shyness, self-focused attention, and behavioral mimicry during social interaction," *J. of Research in Personality*, vol. 98, p. 104225, 2022. 1

[10] D. Roth, D. Mal, C. F. Purps, P. Kullmann, and M. E. Latoschik, "Injecting nonverbal mimicry with hybrid avatar-agent technologies: A naïve approach," in *Proc.\ SUI*, SUI '18, pp. 69–73, 2018. 1, 2, 3, 6, 9

[11] A. Zheleva, J. Hardeman, W. Durnez, C. Vanroelen, J. De Bruyne, D. O. Tutu, J. Morton, J. Joundi, J. Saldien, and K. Bombeke, "The impact of eye gaze on social interactions of females in virtual reality: The mediating role of the uncanniness of avatars and the moderating role of task type," *Heliyon*, vol. 9, no. 10, 2023. 1

[12] T. Wakabayashi, Y. Okada, and K. Zempo, "Effect of salesperson avatar automatically mimicking customer's nodding on the enjoyment of conversation in virtual environments," in *Proc.\ AHs*, AHs '23, pp. 334—-337, 2023. 1, 2, 3, 4, 5

[13] W. Sato and S. Yoshikawa, "Spontaneous facial mimicry in response to dynamic facial expressions," *Cognition*, vol. 104, no. 1, pp. 1–18, 2007. 1

[14] I. Wang and J. Ruiz, "Examining the use of nonverbal communication in virtual agents," *Int. J. of Human–Computer Interaction*, vol. 37, no. 17, pp. 1648–1673, 2021. 1

[15] I. Wang, R. Calvo, H. Wang, and J. Ruiz, "Stop copying me: Evaluating nonverbal mimicry in embodied motivational agents," in *Proc.\ IVA*, IVA '23, pp. 1–4, 2023. 1, 2

[16] N. Aburumman, M. Gillies, J. A. Ward, and A. F. d. C. Hamilton, "Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions," *Int. J. of Human-Computer Studies*, vol. 164, p. 102819, 2022. 1, 2, 3, 6

[17] S. Y. Oh, J. Bailenson, N. Krämer, and B. Li, "Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments," *PLOS ONE*, vol. 11, no. 9, p. e0161794, 2016. 2

[18] J. Hale and A. F. D. C. Hamilton, "Testing the relationship between mimicry, trust and rapport in virtual reality conversations," *Scientific Rep.*, vol. 6, 2016. 2

[19] L. Wen, J. Zhou, W. Huang, and F. Chen, "A survey of facial capture for virtual reality," *IEEE Access*, vol. 10, pp. 6042–6052, 2022. 2

[20] P. Caserman, A. Garcia-Agundez, and S. Göbel, "A survey of full-body motion reconstruction in immersive virtual reality applications," *IEEE Tran. on Visualization and Computer Graphics*, vol. 26, no. 10, pp. 3089–3108, 2019. 2

[21] Meta, "Use Natural Facial Expressions on Meta Quest Pro," 2022. [accessed 14 September 2024]. 2

[22] A. Visconti, D. Calandra, and F. Lamberti, "Comparing technologies for conveying emotions through realistic avatars in virtual reality-based metaverse experiences," *Computer Animation & Virtual Worlds*, vol. 34, no. 3–4, p. e2188, 2023. 2

[23] J. L. Lakin, V. E. Jefferis, C. M. Cheng, and T. L. Chartrand, "The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry," *J. of Nonverbal Behavior*, vol. 27, no. 3, pp. 145–162, 2003. 2

[24] T. Weronika and P. Brzóska, "How behavioral mimicry influences impression formation processes: A positive impression bias," *Polskie Forum Psychologiczne*, vol. 27, no. 1, pp. 5–22, 2022. 2

[25] T. L. Chartrand and J. L. Lakin, "The antecedents and consequences of human behavioral mimicry," *Ann. Rev. of Psychology*, vol. 64, pp. 285–308, 2013. 2

[26] J. L. Lakin and T. L. Chartrand, "Using nonconscious behavioral mimicry to create affiliation and rapport," *Psychological Science*, vol. 14, no. 4, pp. 334–339, 2003. 2

[27] R. B. Van Baaren, R. W. Holland, K. Kawakami, and A. Van Knippenberg, "Mimicry and prosocial behavior," *Psychological Science*, vol. 15, no. 1, pp. 71–74, 2004. 2

[28] W. W. Maddux, E. Mullen, and A. D. Galinsky, "Chameleons bake bigger pies and take bigger pieces: Strategic behavioral mimicry facilitates negotiation outcomes," *J. of Experimental Social Psychology*, vol. 44, no. 2, pp. 461–468, 2008. 2, 9

[29] N. Guéguen, A. Martin, S. Meineri, and J. Simon, "Using mimicry to elicit answers to intimate questions in survey research," *Field methods*, vol. 25, no. 1, pp. 47–57, 2013. 2

[30] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception–behavior link and social interaction.," *J. of Personality and Social Psychology*, vol. 76, no. 6, p. 893, 1999. 2

[31] M. Stel, S. Rispens, M. Leliveld, and A. M. Lokhorst, "The consequences of mimicry for prosocials and proselfs: Effects of social value orientation on the mimicry–liking link," *European J. of Social Psychology*, vol. 41, no. 3, pp. 269–274, 2011. 2

[32] M. Drury and H. van Swol, "Are people who mimic others perceived as more friendly, likeable, persuasive, and knowledgeable?," in *Proc.\ NCA*, 2005. 2, 5

[33] R. B. Van Baaren, R. W. Holland, B. Steenaert, and A. Van Knippenberg, "Mimicry for money: Behavioral consequences of imitation," *J. of Experimental Social Psychology*, vol. 39, no. 4, pp. 393–398, 2003. 2, 5

[34] U. Hess and A. Fischer, "Emotional mimicry as social regulation," *Personality and Social Psychology Rev.*, vol. 17, no. 2, pp. 142–157, 2013. 2, 8

[35] J. N. Bailenson and N. Yee, "Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments," *Psychological Science*, vol. 16, no. 10, pp. 814–819, 2005. 2, 9

[36] E. L. Fasya, E. van den Bos, D. K. Heylen, and M. E. Kret, "Smile mimicry smoothens human-virtual human interactions," *Int. J. of Human-Computer Studies*, vol. 183, pp. 1–11, 2024. 2

[37] F. M. F. Verberne, J. Ham, A. Ponnada, and C. J. H. Midden, "Trusting digital chameleons: The effect of mimicry by a virtual social agent on user trust," in *Proc.\ PERSUASIVE* (S. Berkovsky and J. Freyne, eds.), vol. 7822 of *Lecture Notes in Computer Science*, pp. 234–245, Springer, Berlin, 2013. 2, 9

[38] R. Hoegen, J. van der Schalk, G. Lucas, and J. Gratch, "The impact of agent facial mimicry on social behavior in a prisoner's dilemma," in *Proc.\ IVA)*, IVA '18, pp. 275—-280, 2018. 2

[39] D. S. Choi, J. Park, M. Loeser, and K. Seo, "Improving counseling effectiveness with virtual counselors through nonverbal compassion involving eye contact, facial mimicry, and head-nodding," *Scientific Rep.*, vol. 14, no. 1, pp. 1–12, 2024. 2, 4, 8, 9

[40] M. Suda and M. Oka, "Evaluation of the effect of mimicry on facial expression in avatar-mediated communication," in *Proc.\ ALIFE*, Artificial Life Conference Proceedings, pp. 48–56, 2021. 2, 4, 6, 9

[41] K. Suzuki, M. Yokoyama, S. Yoshida, T. Mochizuki, T. Yamada, T. Narumi, T. Tanikawa, and M. Hirose, "FaceShare: Mirroring with pseudo-smile enriches video chat communications," in *Proc.\ CHI*, CHI '17, pp. 5313—5317, 2017. 3, 8

[42] D. Roth, M. E. Latoschik, K. Vogeley, and G. Bente, "Hybrid avatar-agent technology – a conceptual step towards mediated "social" virtual reality and its respective challenges," *i-com*, vol. 14, no. 2, pp. 107–114, 2015. 3

[43] D. Roth, P. Kullmann, G. Bente, D. Gall, and M. E. Latoschik, "Effects of hybrid and synthetic social gaze in avatar-mediated interactions," in *Proc.\ ISMAR-Adjunct*, pp. 103–108, 2018. 3

[44] D. Roth, C. Klelnbeck, T. Feigl, C. Mutschler, and M. E. Latoschik, "Beyond replication: Augmenting social behaviors in multi-user virtual realities," in *Proc.\ VR*, pp. 215–222, 2018. 3

[45] P. Ekman, "Are there basic emotions?," *Psychological Rev.*, vol. 99, no. 3,

pp. 550–553, 1992. 4

[46] R. Somarathna, T. Bednarz, and G. Mohammadi, “Virtual reality for emotion elicitation–a review,” *IEEE Tran. on Affective Computing*, vol. 14, no. 4, pp. 2626–2645, 2022. 4

[47] F. Huo, R. Gao, C. Sun, and G. Hou, “Age differences in hazard perception of drivers: The roles of emotion,” *Front. in Psychology*, vol. 13, pp. 1–11, 2022. 4

[48] R. Lipson-Smith, J. Bernhardt, E. Zamuner, L. Churilov, N. Busietta, and D. Moratti, “Exploring colour in context using virtual reality: Does a room change how you feel?,” *Virtual Reality*, vol. 25, pp. 631–645, 2021. 4

[49] M. M. Bradley and P. J. Lang, “International affective picture system,” in *Encyclopedia of Personality and Individual Differences* (V. Zeigler-Hill and T. K. Shackelford, eds.), pp. 1–4, Springer Int. Publishing, 2017. 4

[50] Y. Ge, G. Zhao, Y. Zhang, R. J. Houston, and J. Song, “A standardised database of chinese emotional film clips,” *Cognition and Emotion*, vol. 33, no. 5, pp. 976–990, 2019. 4

[51] L. O. H. Kroczek and A. Mühlberger, “Time to smile: How onset asynchronies between reciprocal facial expressions influence the experience of responsiveness of a virtual agent,” *J. of Nonverbal Behavior*, 2023. 4

[52] T. L. Kraft-Feil, R. E. Ingram, C. Gorey, J. H. Luu, M. P. Cross, and S. D. Pressman, “The association of negative mood with automatic and effortful facial expression mimicry,” *Front. in Psychology*, vol. 14, pp. 1–14, 2023. 4

[53] A. Orlowska, M. Rychlowska, and E. Krumhuber, *The Interplay Between Mimicry and Social Context in Facial Expression Perception*, pp. 121–153. Leya, 2020. 4

[54] U. Dimberg, M. Thunberg, and K. Elmehed, “Unconscious facial reactions to emotional facial expressions,” *Psychological Science*, vol. 11, no. 1, pp. 86–89, 2000. 4

[55] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, *Emotional Contagion*. Cambridge University Press, 1994. 4

[56] P. Ekman and W. V. Friesen, “Facial action coding system,” *Environmental Psychology & Nonverbal Behavior*, pp. 1–6, 1978. 4

[57] B. Farnsworth, “Facial Action Coding System (FACS) – A Visual Guidebook - iMotions.” https://imotions.com/blog/learning/research-fundamentals/facial-action-coding-system/. [accessed 04 January 2025]. 4

[58] J. N. Bailenson, N. Yee, K. Patel, and A. C. Beall, “Detecting digital chameleons,” *Computers in Human Behavior*, vol. 24, no. 1, pp. 66–87, 2008. 4

[59] B. Kidwell, V. Lopez-Kidwell, C. Blocker, and E. M. Mas, “Birds of a feather feel together: Emotional ability similarity in consumer interactions,” *J. of Consumer Research*, vol. 47, no. 2, pp. 215–236, 2020. 4, 9

[60] C. Oh, J. Bailenson, and G. Welch, “A systematic review of social presence: Definition, antecedents, and implications,” *Front. in Robotics and AI*, vol. 5, pp. 1–35, 2018. 4

[61] kurtdekker, “Head shake no and nod yes gesture detector · GitHub.” https://gist.github.com/kurtdekker/0f5f353911d7a064e20868e4e5f4dba4. [accessed 04 January 2025]. 4

[62] U. Hess, M. Rychlowska, and T. e. a. Storz, “Bridging the empathy gap: Or not? reactions to ingroup and outgroup facial expressions,” *J. of Cultural Cognitive Science*, 2022. 5

[63] J. Dong, K. Ota, and M. Dong, “User experience of different groups in social vr applications: An empirical study based on user reviews,” *IEEE Trans. on Computational Social Systems*, vol. 11, no. 6, pp. 7133–7145, 2024. 5

[64] H. R. Hatfield, S. J. G. Ahn, M. Klein, and K. L. Nowak, “Confronting whiteness through virtual humans: A review of 20 years of research in prejudice and racial bias using virtual environments,” *J. of Computer-Mediated Communication*, vol. 27, no. 6, 2022. 5

[65] J. R. J. Neo, A. S. Won, and M. M. Shepley, “Designing immersive virtual environments for human behavior research,” *Front. in Virtual Reality*, vol. 2, 2021. 5

[66] S. H. Cha, S. Zhang, and T. W. Kim, “Effects of interior color schemes on emotion, task performance, and heart rate in immersive virtual environments,” *J. of Interior Design*, vol. 45, no. 4, pp. 51–65, 2020. 5

[67] L. Baccon, C. Elodie, and H. MacDougall, “Virtual reality for teletherapy: Avatars may combine the benefits of face-to-face communication with the anonymity of online text-based communication,” *Cyberpsychology, Behavior, and Social Networking*, vol. 22, pp. –, 2018. 5, 6

[68] A. Bradbury, S. Schertz, and E. Wiebe, “How does virtual reality compare? The effects of avatar appearance and medium on self-disclosure,” *Proc.\ HFES*, 2022. 5, 6

[69] M. Gnacek, L. Quintero, I. Mavridou, E. Balaguer-Ballester, T. Kostoulas, C. Nduka, and E. Seiss, “AVDOS-VR: Affective video database with physiological signals and continuous ratings collected remotely in VR,” *Scientific Data*, vol. 11, pp. 1–X, 2024. 5

[70] M. Jeon and W. Zhang, “Sadder but wiser? effects of negative emotions on risk perception, driving performance, and perceived workload,” in *Proc.\ HFES*, vol. 57, pp. 1–X, 2013. 5

[71] C. Kruzic, D. Kruzic, F. Herrera, and J. Bailenson, “Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments,” *Scientific Rep.*, vol. 10, p. 20626, 2020. 6, 9

[72] C. R. Agnew, “Feeling close to a crab-thing in virtual reality: Does avatar appearance always matter in forming meaningful connections? A case study,” *Front. in Virtual Reality*, vol. 3, pp. 1–10, 2022. 6

[73] A. Aron, E. Melinat, E. N. Aron, R. D. Vallone, and R. J. Bator, “The experimental generation of interpersonal closeness: A procedure and some preliminary findings,” *Personality and Social Psychology Bulletin*, 1997. 6

[74] F. Lang, D. John, O. Lüdtke, J. Schupp, and G. Wagner, “Short assessment of the big five: Robust across survey methods except telephone interviewing,” *Behavior Research Methods*, vol. 43, pp. 548–67, 2011. 6

[75] S. Aseeri and V. Interrante, “The influence of avatar representation on interpersonal communication in virtual social environments,” *IEEE Trans. on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2608–2617, 2021. 6, 9

[76] J. C. McCroskey and T. A. McCain, “The measurement of interpersonal attraction,” *Speech Monographs*, vol. 41, no. 6, pp. 261–266, 1974. 6

[77] C. Harms and F. Biocca, “Internal consistency and reliability of the networked minds measure of social presence,” in *Proc.\ PRESENCE*, 2004. 6

[78] D. Davis and W. T. Perkowitz, “Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction.,” *J. of Personality and Social Psychology*, vol. 37, no. 4, p. 534, 1979. 6, 8

[79] M. Burgoon, “Amount of conflicting information in a group discussion and tolerance for ambiguity as predictors of task attractiveness,” *Communications Monographs*, vol. 38, no. 2, pp. 121–124, 1971. 6

[80] J. Kätsyri, K. Förger, M. Mäkäräinen, and T. Takala, “A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness,” *Front. in Psychology*, vol. 6, pp. 1–16, 2015. 6

[81] P. Kullmann, T. Menzel, M. Botsch, and M. E. Latoschik, “An evaluation of other-avatar facial animation methods for social vr,” in *Ext. Abstr.\ CHI EA*, pp. 1–7, 2023. 6

[82] G. F. Reed, F. Lynn, and B. D. Meade, “Use of coefficient of variation in assessing variability of quantitative assays,” *Clinical and Vaccine Immunology*, vol. 9, no. 6, pp. 1235–1239, 2002. 8

[83] J. Hale and A. F. D. C. Hamilton, “Testing the relationship between mimicry, trust and rapport in virtual reality conversations,” *Scientific Rep.*, vol. 6, no. 1, p. 35295, 2016. 8

[84] R. E. Maurer and J. H. Tindall, “Effect of postural congruence on client’s perception of counselor empathy,” *J. of Counseling Psychology*, vol. 30, no. 2, pp. 158–163, 1983. 9

[85] P. A. G. Forbes, X. Pan, and A. F. de C. Hamilton, “Reduced mimicry to virtual reality avatars in autism spectrum disorder,” *Journal of Autism and Developmental Disorders*, vol. 46, no. 12, pp. 3788–3797, 2016. 9

[86] S. A. M. Ali and S. Ali, “The use of non-verbal communication in the classroom,” in *Proc.\ ICFLTAL*, pp. 1096–1099, 2011. 9

[87] D. Hepperle, C. F. Purps, J. Deuchler, and M. Wölfel, “Aspects of visual avatar appearance: Self-representation, display type, and uncanny valley,” *The Visual Computer*, vol. 38, no. 4, pp. 1227–1244, 2022. 9

[88] K. Prehn, C. W. Korn, M. Bajbouj, G. Klann-Delius, W. Menninghaus, A. M. Jacobs, and H. R. Heekeren, “The neural correlates of emotion alignment in social interaction,” *Social Cognitive and Affective Neuroscience*, vol. 10, no. 3, pp. 435–443, 2014. 9

[89] S. Bai, T.-L. Wang, C. Li, A. Venkatesh, T. Simon, C. Cao, G. Schwartz, J. Saragih, Y. Sheikh, and S.-E. Wei, “Universal facial encoding of codec avatars from VR headsets,” *ACM Trans. Graphics*, vol. 43, no. 4, pp. 1–22, 2024. 9