AI-Mediated Augmented Emotion Expression of Avatar for Social Mixed Reality

HyunA Seo
hyuna.seo@hcs.snu.ac.kr
Seoul National University
Korea, Republic of

Youngki Lee
youngkilee@snu.ac.kr
Seoul National University
Korea, Republic of

ABSTRACT

Social Mixed Reality (MR) enables immersive social interactions through avatars. However, users often struggle with emotional expression, especially in multitasking scenarios where they must manage both real-world tasks (e.g., cooking, working, commuting) and virtual social interactions simultaneously. Existing methods for enhancing avatar emotion expression either rely on manual selection, which increases cognitive load, or automated selection, which limits user control. To address this trade-off, we propose the concept of an AI-mediated emotionally expressive avatar that distributes roles between AI and the user. Our system automatically detects users' emotional states, selects appropriate avatar gestures, and allows personalized adjustments to enhance expression. This position paper explores key design considerations for balancing usability and controllability in AI-mediated avatar representations, paving the way for more expressive and intuitive Social MR interactions.

CCS CONCEPTS

• **Human-centered computing** → **Virtual Reality**.

KEYWORDS

Social Mixed Reality, Virtual avatar, Emotion, Body gesture

1 INTRODUCTION

Social Mixed Reality (MR) enables immersive social interaction by inviting friends and family into my physical space through virtual avatars. Rich and natural emotional expression via avatars enhances social presence, interaction satisfaction, and trust. Many commercial MR head-mounted displays (HMDs) have advanced toward more realistic emotional expressions, incorporating embodied technologies such as gaze, facial, and body tracking. Despite these advancements, recent studies indicate that users still struggle to express emotions effectively through avatars. These challenges arise from various factors, including unfamiliarity with the device, awkward avatar control, spatial constraints, multitasking, and even reluctance to move their bodies actively.

This position paper explores **multitasking in everyday Social MR use**, where expressing emotions through avatars becomes challenging. Figure 1 shows the example of our use case. In our scenario, users (1) aim to convey rich emotions via their avatar's body gestures in MR environments but (2) are simultaneously engaged in real-world tasks (e.g., cooking, cleaning, working, commuting) that restrict their physical movement. We identify user challenges in this context and propose high-level design considerations for AI-mediated avatar representations to address these issues.

Figure 1: Example scenarios of multitasking during Social MR. Here, the users want to express their emotions through the virtual avatar's body gestures, but their movements are constrained.

2 BACKGROUND

2.1 Augmented Emotion Expression Methods

In the Human-Computer Interaction (HCI) field, researchers have proposed various methods to enhance or add modalities for avatar-based emotional expression, which we refer to as **augmented emotion expression methods**. For example, Lee et al. [5] explore visualizing biosignals (e.g., heart and breathing rate) alongside avatars to improve emotional communication. Jing et al. [4] introduce visual effects (e.g., color changes, background effects) and sound cues to explicitly convey Social Virtual Reality (VR) users' emotional states. Another common approach is using emojis [1, 7]. Baloup et al.[1] introduced non-isomorphic facial expressions, allowing users to display emotions different from their real ones via emoji-based interfaces. While these methods effectively convey emotions, they lack the natural fluidity of facial and bodily expressions [3], which are essential for social presence. Consequently, they are less suitable for everyday Social MR interactions with close friends and family.

2.2 Trade-Off between Usability and Controllability

An unexplored area in augmented emotion expression methods is **how to enable users to select expressions for their current emotional state**. Existing approaches can be categorized into two mechanisms: *(1) manual selection*: where users explicitly choose expression through a provided interface [1, 7], and *(2) automated selection*: where AI (e.g., emotion detection) captures users' emotion types using sensor data and maps them to pre-defined expressions of avatar [6]. These prior works reveal that the method of selecting an avatar's expression inherently involves a **fundamental trade-off between (1) usability** (e.g., reducing cognitive and physical load) **and (2) controllability over emotional expression**.

Most studies rely on manual selection, allowing users to accurately express their intended emotions [1, 7]. However, in our multitasking scenarios, emotional expression is a secondary task, while social interaction and real-world activities (e.g., work, chores) take priority. Requiring users to constantly recognize their emotional state and manually select an expression imposes a high communication burden, making emotionally rich social interaction difficult.

Automated emotion recognition approaches have been proposed to mitigate this cognitive load and usability challenge. These methods capture biophysical proxies of emotion through sensors (e.g., EEG, pupil dilation) to extract emotional features and render the most explicit avatar expression accordingly [4, 6]. By reducing the load of emotional expression, users can focus more on social interactions and real-world tasks, enhancing overall usability. However this automation completely removes users' control over their emotional expressions. For example, some users may prefer less expressive avatars due to their introverted nature, while others may frequently use emblematic gestures (e.g., a thumbs-up for joy, strong arm movements for angry) and want their avatars to reflect this behavioral pattern. Since automated approaches fail to account for personalized expression preferences and styles, they eventually reduce social presence [1].

3 AI-MEDIATED AUGMENTED EMOTION EXPRESSION OF AVATAR

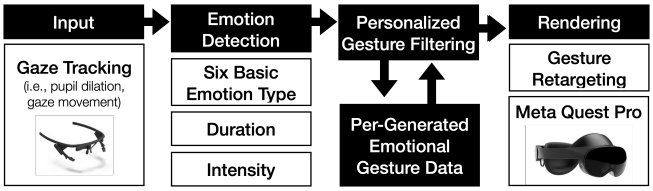

Figure 2: Example of a system pipeline for AI-mediated emotionally expressive avatars.

We propose the concept of an AI-mediated emotionally expressive avatar to facilitate emotional expression in multitasking Social MR scenarios while addressing the usability-controllability tradeoff identified in prior work. A key design consideration for this system is **how to distribute roles between AI and the user** to balance low cognitive load with high controllability in emotion expression. Figure 2 illustrates an example of a system pipeline that integrates our design principles. Our system operates as follows: (1) AI (i.e., emotion detection) identifies the user's current emotional state and automatically selects the most appropriate gesture set, (2) if desired, the user can adjust personalized factors affecting emotional expression to refine the gesture set, and (3) the final expressive gesture is rendered onto the user's avatar. The following sections detail the core design considerations of our system.

3.1 AI: Emotion Detection for Reduced Cognitive Load

As previously discussed, emotion expression in our scenario is a secondary goal. Delegating the emotion recognition and selection of suitable facial expressions and body gestures to AI can significantly reduce user's communication load. Specifically, an emotion detection models can recognize the user's emotion types (e.g., joy, sadness, anger, disgust, fear, surprise, based on Paul Ekman's six universal emotions [2]) and emotion intensities (e.g., low, medium, high). Based on these results, the system selects the most suitable facial expressions and body gestures from a pre-collected animation set.

3.2 User: Personalized Expression for Enhanced Controllability

To address the limitations of automated selection, we introduce user-adjustable controls that allow for personalized fine-tuning of avatar expressions. Users can modify their avatar's emotional expressions in the following ways:

- Adjusting introversion-extroversion levels to control the expressiveness of the avatar.
- Amplifying or softening emotional strength to refine how strongly emotions are conveyed relative to the detected level.
- Adjusting the frequency of emblematic gestures (e.g., thumbs-up, head shake) to match their personal expression habits.

These adjustments can be made at the user's discretion during social interactions through our virtual interface. Factors such as mood (e.g., feeling tired or energetic), personal expression preferences, and social context (e.g., formal vs. casual conversations) may influence how users fine-tune their avatar expressions. This variability highlights why AI cannot generalize and fully automate emotion expression for all users. Once configured, the system dynamically applies the most suitable expressions based on emotion detection results, ensuring a more natural and personalized avatar representation.

4 DISCUSSION

4.1 Allowing User Intervention When AI Prediction is Incorrect

Current emotion detection models achieve an accuracy of around 75–85% [8, 9], which means avatars sometimes express unintended emotions (e.g., stomping in *anger* instead of showing *joy*). Even with advancements in emotion recognition, achieving perfect accuracy is unrealistic. To address this, users should be able to intervene when necessary to ensure correct emotional expression. The challenge is to enable this intervention while keeping the cognitive load low and fostering a mental model that accounts for AI's imperfections without undermining trust. Since little research has explored the integration of uncertain AI models into Social MR interactions, this remains a valuable area for future study.

4.2 Personalization Strategies for Emotion Expression in Social MR

In future work, we aim to explore how user traits and contextual factors shape emotion expression strategies in Social MR. This is a complex question without a clear answer. A common assumption, for example, is that introverted individuals prefer their avatars to reflect their reserved nature. However, given the unique characteristics

[1] Personalized avatar fosters social presence in Social MR/VR.

of virtual avatars, system-assisted expressive behaviors could enable richer and more extroverted social interactions, even for these users. Various factors—such as a user's introversion/extroversion, familiarity with MR devices, usage environment, current mood and energy levels, and the conversational context in Social MR—may all influence how they wish to express emotions. Understanding how these factors shape personalization strategies within our system will offer valuable insights for designing future Social MR applications.

REFERENCES

[1] Marc Baloup, Thomas Pietrzak, Martin Hachet, and Géry Casiez. 2021. Non-isomorphic interaction techniques for controlling avatar facial expressions in vr. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. 1–10.

[2] Paul Ekman, Tim Dalgleish, and M Power. 1999. Basic emotions. *San Francisco, USA* (1999).

[3] Alexander Giovannelli, Jerald Thomas, Logan Lane, Francielly Rodrigues, and Doug A Bowman. 2023. Gestures vs. emojis: Comparing non-verbal reaction visualizations for immersive collaboration. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4772–4781.

[4] Allison Jing, Theophilus Teo, Jeremy McDade, Chenkai Zhang, Yi Wang, Andrei Mitrofan, Rushil Thareja, Heesook Shin, Yongho Lee, Youn-Hee Gil, et al. 2024. Superpowering Emotion Through Multimodal Cues in Collaborative VR. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 160–169.

[5] Sueyoon Lee, Abdallah El Ali, Maarten Wijntjes, and Pablo Cesar. 2022. Understanding and designing avatar biosignal visualizations for social virtual reality entertainment. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[6] Maryam Nadeem, Raza Imam, Rouqaiah Al-Refai, Meriem Chkir, Mohamad Hoda, and Abdulmotaleb El Saddik. 2024. EVOKE: Emotion Enabled Virtual Avatar Mapping Using Optimized Knowledge Distillation. In *2024 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 1–6.

[7] Luyao Shen, Xian Wang, Sijia Li, Lik-Hang Lee, Mingming Fan, and Pan Hui. 2024. EmojiChat: Toward Designing Emoji-Driven Social Interaction in VR Museums. *International Journal of Human–Computer Interaction* (2024), 1–17.

[8] Hao Wu, Jinghao Feng, Xuejin Tian, Edward Sun, Yunxin Liu, Bo Dong, Fengyuan Xu, and Sheng Zhong. 2020. EMO: Real-time emotion recognition from single-eye images for resource-constrained eyewear devices. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*. 448–461.

[9] Zihan Yan, Yufei Wu, Yang Zhang, and Xiang'Anthony' Chen. 2022. Emoglass: an end-to-end ai-enabled wearable platform for enhancing self-awareness of emotional health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.