Do AI-Generated Emoji Overlays from Live Speech Increase Connectedness and Empathy in Professional Video Calls?

AI generated Emojis in video calls

Maha Khalid Alhobishi

King Fahd University of Petroleum & Minerals (KFUPM), g202422100@kfupm.edu.sa

In this project, we are exploring the ability of AI-generated emoji overlay on live speech to make participants feel more connected and empathic when used in a professional video-conferencing session. Studies touched on earlier in the human-computer interaction area have established that facial expressions and gestures are further examples of nonverbal cues which affect the quality of remote interactions in a large scale. Nevertheless, the vast majority of available systems have been operated using a manual trigger, or through face recognition, like in the case of EmojiCam (Namikawa in 2021) [1], and seldom tried in the workplace environment. No research is currently investigating such a speech-driven AI involving listening to live conversation and producing real-time responses of emojis during the course of a video conference. Based on these gaps, the suggested study has two thematic themes: (1) social presence, connectedness, and empathy in professional work involving video-mediated interactions, and (2) speech-based affect recognition and automatic emotional responses in HCI. Through experimental means, the aim of the study is to show whether AI-enhanced interfaces can boost connectedness and empathy among video calls participants compared to standard video call interfaces. Following submissions the methodology and experimental plan will be described.

CCS CONCEPTS

- Human centered computing - Computer-supported cooperative work collaborative and social computing systems and tools.
- Human centered computing - Human computer interaction (HCI) → Empirical studies in HCI.

Keyword

AI- generated Emojis, Video conferencing, social presence, empathy, emojis, sentiment meter, affective computing, HCI

1 INTRODUCTION

Video conferencing has become central to professional collaboration in the 21st century. Yet, participants often note the loss of nonverbal cues—such as micro-expressions and movements—that make remote interactions feel less persona [2]. Human-computer interaction (HCI) researchers have explored ways to restore these signals, with real-time interventions like emoji reactions or sentiment displays showing promise for engagement, though most studies rely on student samples, lab settings, or manual triggers. Nevertheless, these studies are not infallible: they have generally referred to student populations, experimental settings or automated intelligence-controlled systems as opposed to systems relying on manual triggers.

Building on this, the current project investigates whether an AI system that listens to live speech and generates real-time emoji overlays can enhance connectedness and empathy more effectively than standard video interfaces. Unlike prior systems based on facial expression recognition or user-activated cues, this project emphasizes speech-controlled AI for real-time feedback in professional calls, where trust and empathy are critical.

This submission focuses on reviewing and mapping the literature, identifying proxy studies and conceptual gaps. Future phases will detail methodology, participant recruitment, AI system implementation, and measures for evaluating connectedness and empathy.

2 PROXY PAPERQ

The most relevant prior study is Namikawa's *EmojiCam: Emoji-Assisted Video Communication System* (2021) [1] , which explored emoji overlays as nonverbal cues to enrich video calls. EmojiCam enabled participants to place live emoji icons on their streams to show amusement, agreement, or frustration, though the design placed emojis over the face rather than beside it. Controlled experiments compared baseline calls with EmojiCam-augmented ones, using post-call questionnaires, short interviews, and system data. Results showed that participants valued the feature for expressing subtle emotions, reducing silence, and energizing discussions, but some felt that covering faces disrupted natural cues and questioned its appropriateness in serious contexts.

The study's limitations include a student-based lab sample, reliance on manual or facially triggered emojis, and evaluation limited to satisfaction and expressiveness rather than empathy or connectedness. Addressing these gaps, the current research proposes an AI-driven system that generates speech-based emoji overlays beside the speaker's face. Unlike EmojiCam, it will be tested in professional settings, focusing on outcomes like empathy, trust, and connectedness—key factors for effective teamwork.

3 THEMATIC SUMMARY

Theme 1: Social Presence and Empathy in Video-Mediated Communication

Connection and Social Presence between Modalities:

Studies consistently show that the richer the communication channel, the greater the sense of closeness and presence. In 2023, Nowak [2] conducted controlled experiments comparing conversations between friends through text, audio, and video, measuring bonding and perceived closeness with validated scales. The findings revealed that video communication produced significantly more closeness and trust than either text or audio, emphasizing the importance of visual cues for empathy and connection. At the same time, the results suggest that relying solely on text- or audio-based communication is insufficient to build strong connectedness in the workplace.

Psychological Interests and Emotional Responses:

In other words, the role of emotional states in the power of visual cues is pointed out. Providing the results of laboratory-based video meetings and surveys, Seitz in 2024 [3] demonstrated that the high anxiety lessened the perceived connectedness and provided sustained attention which boosted empathy and engagement. On the same note, Saffaryazdi in 2024 [4] used physiological (heart rate), or facial EMG, as well as behavioral coding when demonstrating that emotional cues were inherently changing the dynamics of the conversation, with participants being increasingly attentive and responsive. However, these signals at times aggravated feelings to a disruptive level during delicate interactions.

Attenuated entrapment and Authenticity:

Jelson in 2025 [5] has examined simulated eye contact during interviews, as results suggested that subjects were more favorable and more authentic to interact more in comparison with a standard condition. However, a number of participants said that the use of artificial cues was unnatural and distracting until it compromised their authenticity. This discrepancy is an indication that as long as cues are enriched, they are able to create more trust but excessive intrusion or artificiality will turn counter productive.

Synthesis:

Researchers consistently show that enriched nonverbal cues enhance empathy, connectedness, and trust, while artificial or intrusive ones risk reducing genuineness. This balance justifies testing AI-generated emoji overlays placed near—but not covering—the speaker's face, keeping them expressive yet non-disruptive. Across methods such as controlled experiments, surveys, physiological monitoring, and prototype testing, findings confirm that richer signals foster social connection. Still, cues that are too obtrusive can backfire, highlighting why side-positioned, speech-triggered emojis may provide a stronger yet less disruptive form of feedback.

Theme 2: Affective Interfaces to Real Time Emotions Cues

Emotional Visualization & Prototype Tools:

Prototypical systems show how emotional feedback has the potential to enhance user involvement. Zhu (2020) proposed a visualization tool EMVAS to show the consistency of emotions when presenting the video [6]. Assessed by the analysis of videos and surveys by the audience, EMVAS made the recognition of the emotional flow better and made the engagement more effective. Nonetheless, since it was experimented on formatted presentations, it is unknown how these systems need to work in interactive, active video calls.

Sentiment detection and Datasets and Algorithmic Advances:

The current literature has highlighted the use of multimodal datasets in enhancing the sentiment recognition. Liu et al. in 2022 [7] created CH-SIMS v2.0, a multimodal dataset consisting of acoustic and visual symptoms to compare acoustic and visual apparatus across algorithms. Multimodal approaches establish other options which although proven highly accurate and made by models trained on this dataset showed high accuracy in the laboratory setting. However, the algorithms actually worked less in toto work under real life dialogue wherein noise and uncertainty made predictions less accurate.

Affective compute Systematic Reviews:

Lin and Zhang in 2022 [8] examined the research conducted on affective computing in the human-computer interaction research process in terms of methods, datasets and applications. They concluded that affect recognition technologies demonstrated stable potential of enhancing engagement, and evaluation of the shortcomings of methodological disaggregation and application in real-time tools persist as a critical issue. Majority of those systems simply continue after detection and do not offer direct feedback to the users.

Synthesis:

Overall, prior studies show that affective interfaces can enhance interaction in certain contexts, but most lack the ability to deliver dynamic, real-time feedback. This highlights the novelty of the proposed study: testing whether AI-based, speech-activated emoji overlays in professional video calls can transform recognition into meaningful cues that foster empathy and connectedness. Research confirms that automatic affect recognition is technically feasible and can improve engagement, with tools like EMVAS and datasets like CH-SIMS achieving high accuracy in lab settings. Yet, most systems stop at recognition without generating actionable feedback for real users. This creates a critical gap: no system has yet deployed AI-driven, speech-based emoji overlays in real-time professional meetings. The present study directly addresses this by evaluating their potential to strengthen empathy and connectedness in workplace communication.

4 RESEARCH GAP

Although previous literature demonstrates the enhancement of medicated communication through emotional cues, there are three gaps to be considered: (1) the majority of studies are lab or educational-based scenarios, instead of the work environment; (2) previous research has not arrived at speech-driven AI systems that autonomously produce emojis, had a reliance (rather) on manual or facial-expression states; and (3) connectedness and empathy is uncommonly a metric. This is dealt with in my proposed work, which spreads EmojiCam to professional environments, with the difference that the AI began to listen to speech and automatically place emojis to determine the first controlled experiment of whether impolicies increased interconnectedness and empathy over a normal interface.

REFERENCES

[1] K. S. I. I. R. S. S. &. O. Y. Namikawa, "EmojiCam: Emoji-Assisted Video Communication System Leveraging Facial Expressions.," In International Conference on Human-Computer Interaction. Cham: Springer International Publishing., pp. 611-625, 2021.

[2] J. e. a. Seitz, " The impact of video meeting systems on psychological user states.," International Journal of Human–Computer Interaction, vol. 40, no. 6, p. 587–603, 2024.

[3] K. T. L. T. J. &. R. S. Nowak, "Hear we are: Spatial audio benefits perceptions of turn-taking and social presence in video meetings.," In Proceedings of the 2nd annual meeting of the symposium on human-computer interaction for work., pp. 1-10, 2023.

[4] K. R. R. M. A. N. M. J. &. F. T. Chow, "Exploring a Real-time Feedback Display of Non-verbal Cues in Online Work Meetings to Support Self-Presentation.," Proceedings of the ACM on Human-Computer Interaction, , vol. 9, no. 2, pp. 1-28, 2025.

[5] X. F. H. C. E. H. Y. J. M. Y. H. &. W. R. Zhu, " EMVAS: end-to-end multimodal emotion visualization analysis system.," Complex & Intelligent Systems, 11(8), 1-15., 2025.

[6] Y. Y. Z. M. H. L. Z. Y. W. Q. Y. .. &. G. K. Liu, "Make acoustic and visual cues matter: Ch-sims v2. 0 dataset and av-mixup consistent module.," In Proceedings of the 2022 international conference on multimodal interaction, pp. 247-258, 2022.

[7] W. &. Z. Y. Lin, "Review of researches on the emotion recognition and affective computing based on HCI.," In 2022 7th International Conference on Image, Vision and Computing (ICIVC).IEEE., pp. 573-578, 2022.

[8] S. e. Z. S. R. U. H. &. N. M. Javed, "How AI can detect Emotional cues in Students, Improving Virtual Learning Environments by providing Personalized support and Enhancing Social-emotional Learning.," Review of Applied Management and Social Sciences, vol. 8, no. 2, pp. 665-682, 2025.

[9] N. K. N. L. G. L. K. B. E. &. B. M. Saffaryazdi, "Exploring the impact of computer-mediated emotional interactions on human facial and physiological responses.," Telematics and Informatics Reports, vol. 14, p. 100131, 2024.

[10] A. T. M. T. L. S. I. K. S. &. L. S. W. Jelson, "Investigating the Effects of Simulated Eye Contact in Video Call Interviews.," In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, pp. 1-12, 2025.