Speejis: Expressive Speech Emotion Cues in Voice Messaging

Ilhan Aslan
Aalborg University
Aalborg, Denmark
ilas@cs.aau.dk

Carla F. Griggio
Aalborg University
Copenhagen, Denmark
cfg@cs.aau.dk

Henning Pohl
Aalborg University
Aalborg, Denmark
henning@cs.aau.dk

Timothy Merritt
Aalborg University
Aalborg, Denmark
merritt@cs.aau.dk

Niels van Berkel
Aalborg University
Aalborg, Denmark
nielsvanberkel@cs.aau.dk

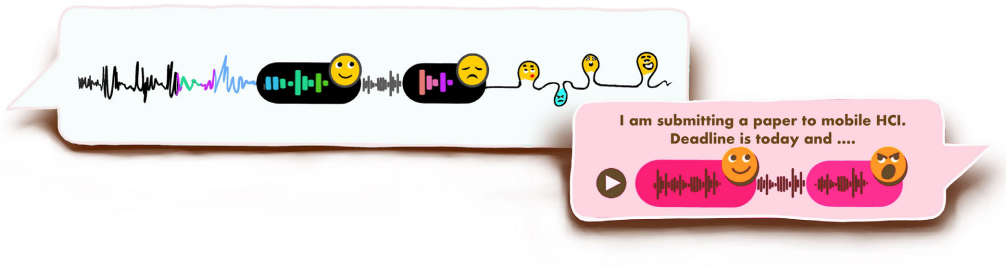


Figure 1: Illustrations of future design directions for voice messaging. This design mockup proactively maps automated speech emotion recognition onto visual cues like emotional waveforms and emojis.

Abstract

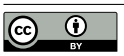
Voice messaging remains a unimodal experience centered on recording and listening to audio. Current speech analysis models could supplement messages with automated transcriptions and visual emotion cues, such as emojis based on the speaker's tone or the verbal content. However, how these proactive features affect UX is not yet well understood. To address this gap, we introduce Speejis, a proactive agent that leverages a Speech Emotion Recognition (SER) model to automatically augment voice messages with emojis and colored waveforms, alongside automatic speech recognition for transcriptions. We present the Speejis prototype and report findings from an exploratory user study (N = 12). Our results detail how Speejis impacts UX, highlighting a trade-off between increased stimulation and reduced dependability due to the fully automated nature of emotion cueing.

CCS Concepts

• **Human-centered computing** → **Interaction design**; **Ubiquitous and mobile computing systems and tools**;

Keywords

Affective Computing, Speech Emotion Recognition, User Experience, Proactive AI, Voice messaging



This work is licensed under a Creative Commons Attribution 4.0 International License.
MobileHCI '26, Swansea, United Kingdom
© 2026 Copyright held by the owner/author(s).
ACM ISBN /26/08
https://doi.org/10.1145/3821581.3833104

ACM Reference Format:
Ilhan Aslan, Carla F. Griggio, Henning Pohl, Timothy Merritt, and Niels van Berkel. 2026. Speejis: Expressive Speech Emotion Cues in Voice Messaging. In *28th International Conference on Mobile Human-Computer Interaction (MobileHCI '26), August 31–September 03, 2026, Swansea, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3821581.3833104

1 Introduction

The integration of emotion detection in digital messaging has been a long-standing effort in HCI. Today's speech analysis models based on deep learning can recognise emotions increasingly well [22, 23]. Consequently, the design space for affective voice messaging is expanding and creating new opportunities to improve user experiences (UX) [12] via AI agents. The need for new experience designs in voice messaging has been recently demonstrated by, for example, An et al. [4]. They have explored the use of emotional teasers when sending a message on a smartwatch; i.e., chat-balloon animations [2], which on the receiving side can serve as enjoyable previews to an incoming voice message, allowing a glimpse at the message context without having to listen to the voice message right away. Broadly, research on messaging demonstrates a long-standing interest in exploring new forms of rich, non-verbal means of expression beyond GIFs, stickers and emoji [6], for example, by allowing manual or automated customizations to a profile picture [26], shape [5] and colour [13] of message bubbles, keyboard colour themes [9], animations [1] and vibration patterns associated to an emoji [3], or combinations of augmentations [19]. However, most of these augmentations have been applied to visual content, and we are interested in the emerging design space of augmented voice messages [10, 11, 15], whose rich non-verbal cues often remain obscured in predominantly visual messaging interfaces.

We believe that messaging users could also benefit from AI taking a more proactive role in augmenting voice message content. This could include generating message transcriptions, or visual content such as emojis and novel audio-waveform designs based on speech analysis. However, the use of automation in design can be controversial and a double-edged sword resulting in potentially obnoxious interaction designs [14].

Overall, the mixed design space of affective speech AI [7] and voice messaging [24] presents unexplored opportunities for proactive designs (e.g., Figure 1), which lead to the overarching research question of this paper: How does a proactive agent that augments voice messages with visual speech emotion cues affect different aspects of user experience? To this end, the paper presents artifact-driven research by introducing Speejis, an experimental prototype of a proactive agent. We report on a user study in a lab setting evaluating initial designs of agent-generated speech emotion cues. In summary, the paper contributes: 1) A replicable architecture to realize proactive and agent-generated speech emotion cues for voice messaging alongside transcriptions; 2) Results of an explorative study (N = 12) with the Speejis prototype, detailing the UX of using Speejis; 3) A discussion of future design opportunities for voice messaging with proactive and agent-generated speech emotion cues.

2 Speejis Prototype

Figure 2 shows an overview of the Speejis system architecture, which is used to process the audio stream from a microphone (**1**). Details of how the Speech Emotion Recognition (SER) model is used are depicted in (**1**) and (**2**). In Speejis the SER model (**2**) provides emotion values of valence, arousal, and dominance for both the smaller audio chunks while the user is speaking and the whole message at the end. The whole message is also saved as an audio file, which is processed by a transcription model (**6**). To perform SER, Speejis uses the model provided by Wagner et al. [23]. To transcribe the voice message into text, it uses the base Whisper model [21]. The two aforementioned mapping models are applied to map emotion labels to emojis (**4**) and waveform colours (**3**), resulting in the design material (**5**), which includes the message transcript(**6**) and is used to compose voice messages with non-verbal cues. Considering latency and computational requirements, we tested the system on a MacBook Air (M4) with and without using the Whisper model for message transcription. Without the Whisper model, the inference time needed for speech emotion recognition was barely noticeable. When using the system with the additional Whisper model, a slight delay was introduced, depending on the voice message length, of about 1-3 seconds, which we deemed as acceptable. Consequently, we decided to include text transcription as a feature, both as an alternative to use with or without visual emotion cues. Figure 3a illustrates an example voice message and all options baseline and new designs for the UI.

The Speejis agent makes use of the emoji classifications provided by Kutsuzawa et al. [16], who provide valence and arousal values for 74 facial emojis. For this explorative study, we selected 22 of these emojis (see Figure 3b) representing a broad range of emotional expressions.

In addition to using emojis, we explored ways to also augment the audio waveform, which is typically shown as part of a voice message. We created a first version of an emotional audio waveform using a simple colour mapping (see Figure 3c) to indicate valence and arousal for individual bars in the audio waveform.

To study the effects of voice messaging with proactive and agent-generated speech emotion cues, we implemented a web application. The Flask Python web framework is used as the backbone, delivering web-based user interfaces that can be customized for mobile usage, which is the most popular device for digital and asynchronous voice messaging. Consequently, the UI is based on web technologies, specifically HTML emojis and CSS.

3 User Study

We conducted an exploratory user study to gather preliminary insights into the user experience of voice messaging with Speejis. We were interested in identifying potential benefits and limitations to inform the design of proactive and agent-generated emotion cues. We also sought to gather higher-level user perspectives on the use of proactive AI to generate emotion cues in voice messaging.

3.1 Apparatus

An iPhone 12 Pro was used as the front-end to experience voice messaging, while the Speejis system was run on a MacBook Air (M4). The iPhone and MacBook Air were connected via a dedicated WiFi access point to enable communication for all back-end processing. A desktop condenser microphone was used to both stream the voice message and record the posthoc interviews with the participants. We planned 30 minutes for each session, with some participants taking a few minutes longer to discuss the potential implications of using SER in voice messaging. We seated the participants next to the desktop condenser microphone and let them believe that the recording was happening on the mobile. When participants played the voice message, the audio of the recording was played on the iPhone.

3.2 Participants

We recruited 12 (eight female, four male), students (six) and staff members from the University campus between the ages of 25-44 years (M=29), with varying experiences in using voice messaging in their daily lives.

3.3 Procedure

Figure 3d illustrates the setting for the user study. The first six participants took part in the study alone. For the next six participants we decided to test as "pairs", participants still tested the system one after the other and had to answer the structured questions individually, but the discussion at the end was together. This allowed for facilitating deeper discussions among pairs who know and communicate regularly with each other. The general procedure was welcoming them, explaining the basic concept of voice messaging with agent-generated speech emotion cues, demonstrating the system, and showing them examples of voice messages with and without the Speejis agent. A printed version of the designs were provided to the participants as a reference of which (groups of) designs should be compared in the study.

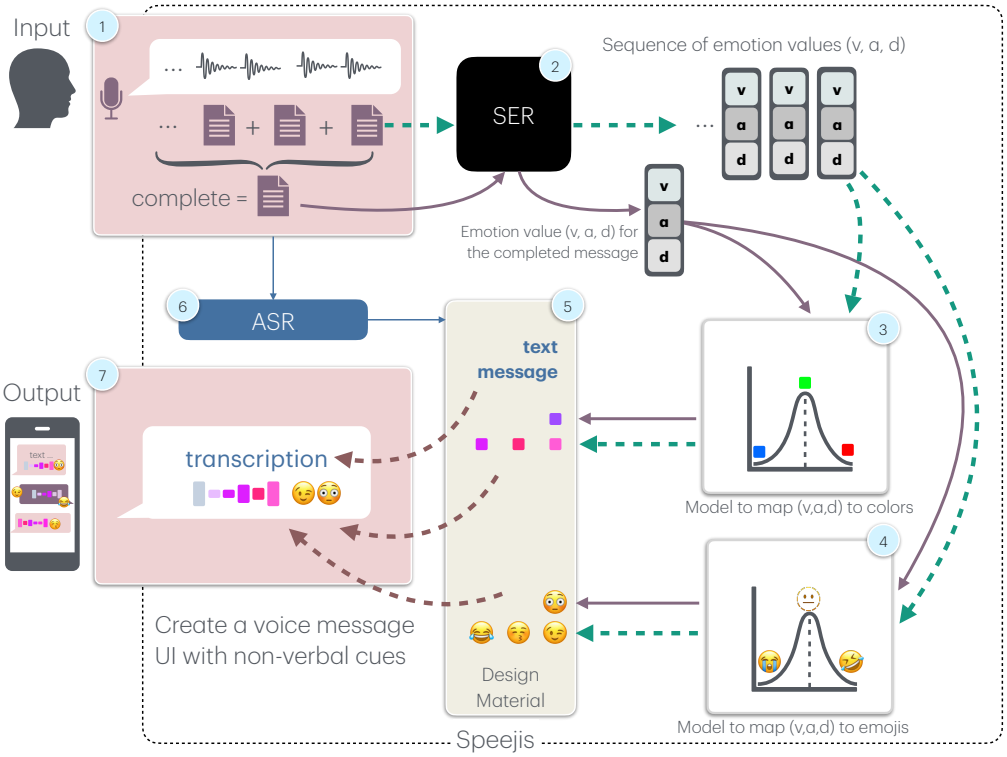


Figure 2: Overview of the Speejis' system architecture illustrating the components of the system and how they interact with each other to provide the material needed to proactively augment voice messages

The participants then used the system for as long as they wanted. Having pairs of participants was beneficial as participants not only saw the results of their own voice messages but also witnessed another person's usage and results before entering the discussion. We instructed participants to leave any voice message they wanted to test how the Speejis agent would incorporate non-verbal emotional cues and assess the accuracy of the transcription. Participants could leave a voice message recording and see resulting designs (see Figure 3a), one below the other and mixed in order. Considering the explorative nature of the study, we did not randomize the order after every new message, which could be seen as a limitation causing an order effect. However, participants could scroll up and down through all of the displayed voice message design alternatives for as long as they wanted. Through buttons placed after the last design they could listen to an audio recording of their voice message or record a new message. All participants were encouraged to explore the system and try leaving messages with different intended feelings or tones. Participants were also invited to record mixed emotions in the message to transition the emotional tones in the same message to challenge the systems capabilities, e.g., starting with a happier tone and ending on a sadder tone.

When participants were ready to provide feedback, they were first asked to fill out the English version of the UEQ [17], both for the voice messaging experience with and without Speejis. Afterwards, we conducted a semi-structured interview during which we encouraged participants to revisit the prototype and test it again, to reflect on their comments, or demonstrate to us what they were referring to for clarification. We started the interview by asking if they would use Speejis in their personal communication, if they answered yes we asked, which version of the design (i.e., with emojis, coloured waveform, or combinations) they would prefer and why. We asked them to explain what they liked and disliked, and whether their answers or preferences would change if they were on the receiving end of the message. We also asked if they had any suggestions for improving future design iterations. The rest of the interview was unstructured and invited discussion about how they felt about the level of automation, and any other topics they would want to share with us regarding their experience.

4 Results

We analysed user experience through a mixed methods approach.

4.1 Analysis of the UEQ Data

As each participant rated the design with and without Speejis, providing UEQ ratings for both solutions, we ran a paired-samples t-test for each dimension of the UEQ. The handout for the participants included a printed version of the baseline (i.e. 1. a design with only a black waveform and 2. the black waveform in combination with the message transcription) and Speejis designs (all designs with emojis or coloured waveforms). The reported p-values have been adjusted with Holm-Bonferroni correction (adjusted to address the use of multiple t-tests).

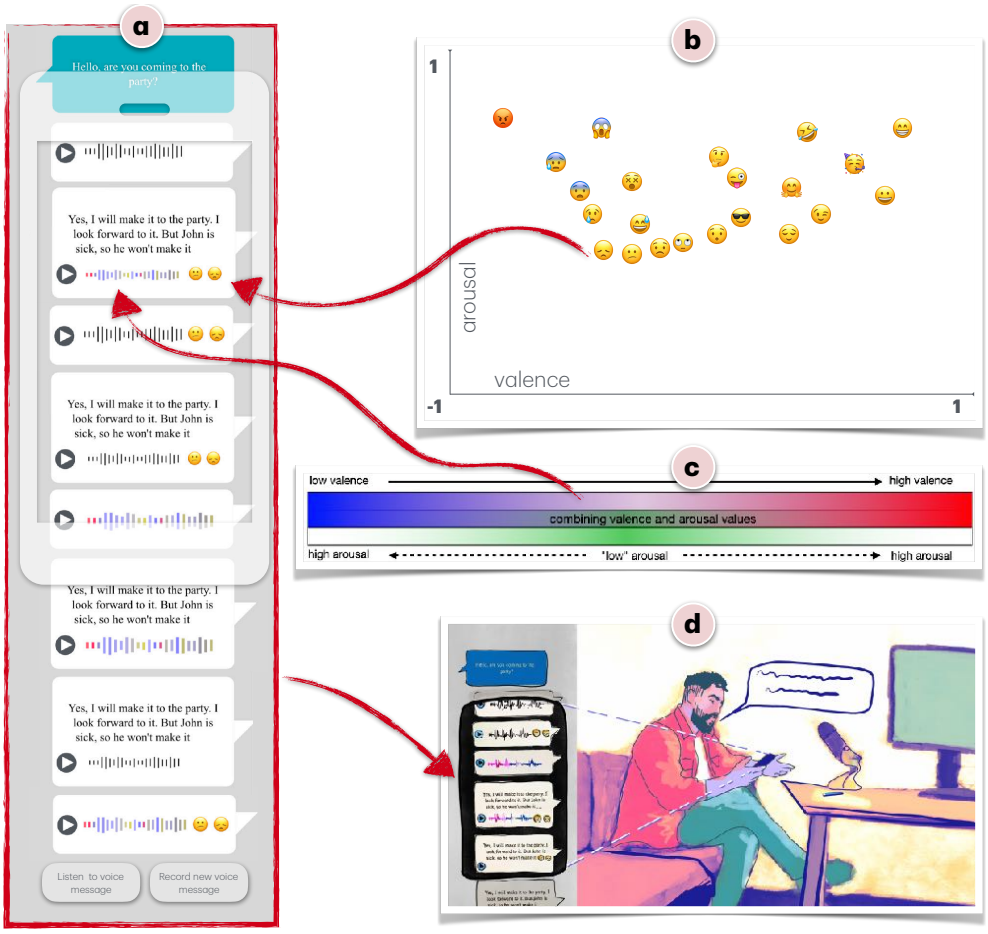


Figure 3: a) Example of design probes used in the study, include the baseline and new conditions both with and without a transcription. b) The set of 22 emojis used by the Speejis agent in the study to automatically augment voice messages with emojis. c) Concept for the colour mapping used by Speejis to augment the waveform. d) An illustration of the lab study setup.

We found that use of Speejis had a significant main effect on ratings of attractiveness ($p = .008$, $t(11) = 4.01$, $d = 1.16$), novelty($p = .001$, $t(11) = 5.54$, $d = 1.60$)), and stimulation($p = .005$, $t(11) = 4.42$, $d = 1.28$). Ratings for these constructs were higher for designs with Speejis. In Figure 4a, we can see that the difference in mean ratings for these constructs is considerable. While the difference in novelty may not be a surprise, the high difference in stimulation and attractiveness indicates why participants overall preferred to use Speejis in voice messaging. There also was a significant effect on dependability ($p = .029$, $t(11) = -3.12$, $d = -0.90$), with participants rating the designs with Speejis as (slightly) less dependable. As also shown in Figure 4a, while the difference in dependability is significant, both conditions received positive ratings.

Figure 4b provides user ratings for each item of the UEQ questionnaires, providing details of how users rated the UX of voice messaging with vs without. For example, we can see that the Speejis UX is rated more enjoyable, exciting, or friendly than the status quo. We can also see that the proactive design falls short in a few important item ratings, which should be carefully addressed by making the output more predictable and secure in future versions.

4.2 Participants' preferences, likes, and dislikes

For the qualitative analysis of the interviews, we iteratively created and refined mind maps of participant quotes and patterns in the topics of their responses.

All (12) participants reported that they would want to use voice messaging with Speejis and the transcription function, as well as including emojis in the design. Eight preferred the combination of emojis with coloured waveforms, arguing, for example, that the waveform could be used to skip to parts of the message and that the emoji described the emotions well. The other four participants who preferred to use emojis with the standard waveform argued that the colours were adding too much complexity; and that the colour mappings were not intuitively interpretable. One of them argued that the coloured waveform had a "too quantifying feeling to it…" and was also "too jumpy and not really aesthetic".

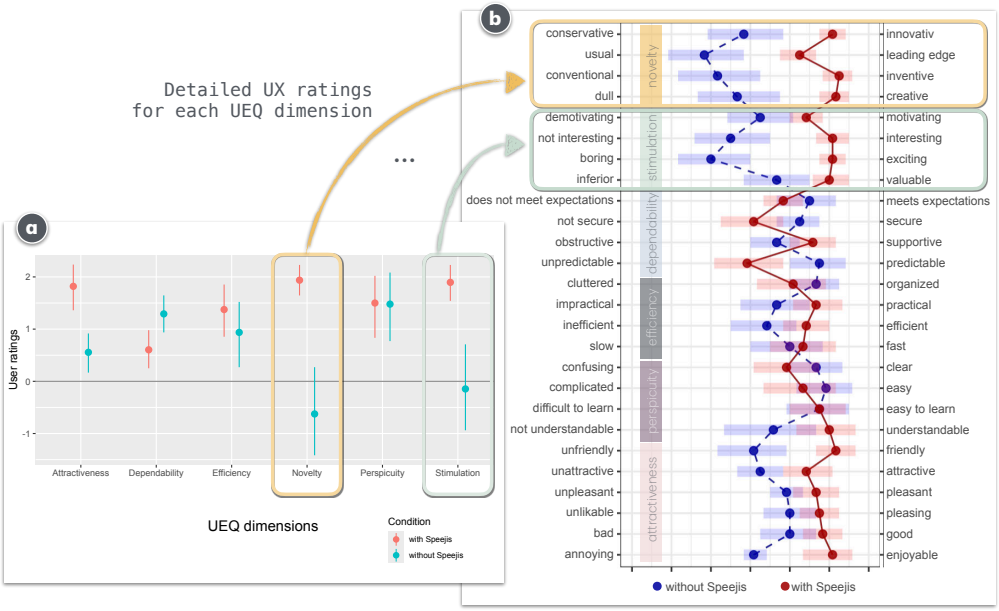


Figure 4: a) Results of the UX questionnaire comparing voice messaging experience with and without visual speech emotion cues. Error bars denote 95% confidence intervals. b) Details of the UEQ items. Error bars denote 95% confidence intervals.

4.2.1 Participants' suggestions on design features. Emojis were described as comforting and recognizable. Two participants mentioned that emojis have some room for ambiguity, which they actually liked. We also asked if the participants would have preferred more or fewer emojis. There was no agreement on the number of emojis and on which parts of the message emojis should represent. However, some participants highlighted that a **longer message may require more emojis** and that **maybe emojis should then only be applied to regions with "extreme" emotions**. Most participants appreciated the fact that the speech emotion recognition was using how participants sounded in the message. However, four participants brought up the topic of also including linguistic analysis for the selection of the emojis, and one participant suggested as an example that when the content of the message is about horse riding, then an emoji "riding a horse" could be used instead of a generic emoji.

Two participants mentioned animations as an alternative to enhance the coloured waveform, such as having each bar move according to its emotion labels, which would add to the liveliness of the design.

4.2.2 What did participants like? Participants liked that Speejis were considered as fun, nice, supportive, helpful, providing an added level of communication, making it more lively, and much **quicker and easier to get a sense of what is going on**, to mention a few of the formulations participants used to describe what they liked about emojis. These comments align with the results of the UEQ questionnaire, where participants rated the overall attractiveness, novelty, and stimulation dimensions of voice messages with Speejis significantly higher than those without Speejis. One participant stated that *"Initially, my reflection was this introduces like an extra sort of layer of complexity. but then it makes it more efficient."*, continuing with examples of how difficult it is sometimes to convey the tone of the message including the example of when they write their partner *"I am on my way home (hot emoji)"* means something completely different or when their father sends them a message stating "call me back" that that can sound like a threat to someone who does not know their father. Another participant highlighted the fact that **voice messages are used when things are more serious** and whenever they get a voice message they are extremely anxious about the content, having **a smiley attached to the message gives them a lot of comfort**, especially when you are not in a situation where you can listen to the voice message right away. These comments suggest that Speejis are also associated with functional qualities and potentially the efficiency dimension, as measured by the UEQ questionnaire.

4.2.3 What did participants not like? Most of the comments referring to what participants did not like about Speejis focused on the **added complexity, mainly associated with the coloured waveform design** but also concerns related to the fact that the augmentation was happening fully automatically, which is reflected for example by the following statement of one of our participants "***it is a little hypothetical, but just the idea that if it got your intention wrong*** *and maybe it was like a really sad message, but you were like, I don't know, nervously laughing throughout and it misunderstood that, it could create potentially uncomfortable situations.*" Consequently, multiple participants stated **editorial power would be good in any case, such as removing emojis or adding one more emoji to the existing were mentioned as examples**. To clarify, none of the participants stated that the created emojis were useless or wrong. One participant stated that if the option (to select manually) is not the default, they probably would not bother to search for it in the settings, but use whatever the default is.

We also asked if marking the emojis as AI-generated would help limit their concern and participants agreed that this idea would indeed remove their concerns. Consequently, an important takeaway is to **make it clear who added the emojis (user or AI)**. It is worth mentioning that some of the participants who mentioned some concerns about the coloured waveform would, however, be more interested in receiving such detailed analysis as the receiver of voice messages. When asked why, one participant explained that it is more important to them to better understand the tones in messages others send to them than the other way around. When it was stated that this processing could happen automatically on the receiving side and how they would feel about this, there was a consensus that the sender should be in control if possible, but there was also some reflection on the fact that such things were out of their control, and if it happened on the receiving device automatically, it would also be fine. However, two participants highlighted that while they would use these features in personal messaging they would not use them in some social network apps.

5 Conclusion and Future Directions for Design

This paper explored the inclusion of visual emotion cues to voice messages. We argued that speech analysis models are increasingly capable and are starting to provide the basic functionality to add visual speech emotion cues fully automatically to voice messages. As such a messaging design shifts some control from the human user to an agent, the design and its potential effect on UX had to be approached critically. To this end, we created the experimental Speejis agent and conducted a user study exploring its effect on UX when Speejis is part of a voice messaging UI. While the study's scope was limited in sample size, it suggests a preference for the proactive design, as it mitigated accessibility challenges and significantly increased stimulation in comparison to the baseline condition without emotional cueing. Critical areas for future work include reducing AI dependency and developing effective error-recovery mechanisms.

Moreover, there are costs associated with dependency on an agent that takes initiative and co-creates the messaging experience. Participants reflected on negotiating their new role in communication, where editorial power, emotion interpretation, and data control are shared between humans and AI agents (as potentially different agents can be used at the receiving and sending end). The humans expose more (or potentially too much) of themselves for the benefit of an "improved" user experience, in terms of stimulation, convenience, and fast communication. However, there is also a risk that emotions could be interpreted as something clear-cut and deterministic, while in human communication, emotions are also helpful when there is room for ambiguous interpretations. Emojis seem to carry this quality in contrast to the colored waveforms deployed in the user study, echoing previous work on emoji ambiguity and appropriation [8, 18, 20, 25]. Future work should prioritize personalization or adaptation, allowing for co-creation of emotion communication between humans and AI, as well as co-customizations [9] of voice augmentations between communication partners. Future studies may also investigate the effects of long-term adoption of co-created augmentations with proactive AI, such as Speejis, studying the extent to which users adapt their communication to AI's suggestions, for example, if users started to deliberately self-style their voice to make the best use of a Speejis agent.

Looking forward, we also believe it is essential to acknowledge that affective computing technologies need to be applied carefully, as they risk being used for profiling users' attitudes and preferences, which may not be in their long-term interest. We encourage future work on protecting users from such "moments of affective vulnerability", which could become a feature in future proactive agents. We hope our research can contribute to the evolution of messaging practices, making future multimodal messaging more expressive and inclusive.

References

[1] Jessalyn Alvina, Chengcheng Qu, Joanna McGrenere, and Wendy E. Mackay. 2019. MojiBoard: Generating Parametric Emojis with Gesture Keyboards. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312771

[2] Pengcheng An, Chaoyu Zhang, Haichen Gao, Ziqi Zhou, Yage Xiao, and Jian Zhao. 2025. AniBalloons: Animated chat balloons as affective augmentation for social messaging and chatbot interaction. *International Journal of Human-Computer Studies* 194 (2025), 103365. https://doi.org/10.1016/j.ijhcs.2024.103365

[3] Pengcheng An, Ziqi Zhou, Qing Liu, Yifei Yin, Linghao Du, Da-Yuan Huang, and Jian Zhao. 2022. VibEmoji: Exploring User-authoring Multi-modal Emoticons in Social Communication. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 493, 17 pages. https://doi.org/10.1145/3491102.3501940

[4] Pengcheng An, Jiawen Stefanie Zhu, Zibo Zhang, Yifei Yin, Qingyuan Ma, Che Yan, Linghao Du, and Jian Zhao. 2024. EmoWear: Exploring Emotional Teasers for Voice Message Interaction on Smartwatches. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 279, 16 pages. https://doi.org/10.1145/3613904.3642101

[5] Toshiki Aoki, Rintaro Chujo, Katsufumi Matsui, Saemi Choi, and Ari Hautasaari. 2022. EmoBalloon - Conveying Emotional Arousal in Text Chats with Speech Balloons. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 527, 16 pages. https://doi.org/10.1145/3491102.3501920

[6] Daniel Buschek, Mariam Hassib, and Florian Alt. 2018. Personal Mobile Messaging in Context: Chat Augmentations for Expressiveness and Awareness. *ACM Trans. Comput.-Hum. Interact.* 25, 4, Article 23 (Aug. 2018), 33 pages. https://doi.org/10.1145/3201404

[7] Caluã de Lacerda Pataca, Saad Hassan, Nathan Tinker, Roshan Lalintha Peiris, and Matt Huenerfauth. 2024. Caption Royale: Exploring the Design Space of Affective Captions from the Perspective of Deaf and Hard-of-Hearing Individuals. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 899, 17 pages. https://doi.org/10.1145/3613904.3642258

[8] Carla F. Griggio, Benjamin M. Gorman, and Garreth W. Tigwell. 2024. Party Face Congratulations! Exploring Design Ideas to Help Sighted Users with Emoji Accessibility when Messaging with Screen Reader Users. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 175 (April 2024), 31 pages. https://doi.org/10.1145/3641014

[9] Carla F. Griggio, Arissa J. Sato, Wendy E. Mackay, and Koji Yatani. 2021. Mediating Intimacy with DearBoard: a Co-Customizable Keyboard for Everyday Messaging. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 342, 16 pages. https://doi.org/10.1145/3411764.3445757

[10] Gabriel Haas, Jan Gugenheimer, Jan Ole Rixen, Florian Schaub, and Enrico Rukzio. 2020. "They Like to Hear My Voice": Exploring Usage Behavior in Speech-Based Mobile Instant Messaging. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 35, 10 pages. https://doi.org/10.1145/3379503.3403561

[11] Gabriel Haas, Jan Gugenheimer, and Enrico Rukzio. 2020. VoiceMessage++: Augmented Voice Recordings for Mobile Instant Messaging. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 30, 10 pages. https://doi.org/10.1145/3379503.3403560

[12] Marc Hassenzahl and Noam Tractinsky. 2006. User experience—a research agenda. *Behaviour & Information Technology* 25, 2 (2006), 91–97. https://doi.org/10.1080/01449290500330331

[13] Mariam Hassib, Daniel Buschek, Paweł W. Wozniak, and Florian Alt. 2017. HeartChat: Heart Rate Augmented Mobile Chat to Support Empathy and Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2239–2251. https://doi.org/10.1145/3025453.3025758

[14] Wendy Ju and Larry Leifer. 2008. The Design of Implicit Interactions: Making Interactive Systems Less Obnoxious. *Design Issues* 24, 3 (07 2008), 72–84. https://doi.org/10.1162/desi.2008.24.3.72

[15] JooYeong Kim, SooYeon Ahn, Yoonjae Kim, and Jin-Hyuk Hong. 2026. Voice-Visualized Message Interactions on Smartwatches. *International Journal of Human–Computer Interaction* in press (March 2026), 28 pages. https://doi.org/10.1080/10447318.2026.2643343

[16] Gaku Kutsuzawa, Hiroyuki Umemura, Koichiro Eto, and Yoshiyuki Kobayashi. 2022. Classification of 74 facial emoji's emotional states on the valence-arousal axes. *Scientific Reports* 12, 1 (2022), 398. https://doi.org/10.1038/s41598-021-04357-7

[17] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.

[18] Hannah Miller, Daniel Kluver, Jacob Thebault-Spieker, Loren Terveen, and Brent Hecht. 2017. Understanding Emoji Ambiguity in Context: The Role of Text in Emoji-Related Miscommunication. In *Proceedings of the International AAAI Conference on Web and Social Media*. 152–161. https://doi.org/10.1609/icwsm.v11i1.14901

[19] Romina Poguntke, Tamara Mantz, Mariam Hassib, Albrecht Schmidt, and Stefan Schneegass. 2019. Smile to Me: Investigating Emotions and their Representation in Text-based Messaging in the Wild. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC '19)*. Association for Computing Machinery, New York, NY, USA, 373–385. https://doi.org/10.1145/3340764.3340795

[20] Henning Pohl, Christian Domin, and Michael Rohs. 2017. Beyond Just Text: Semantic Emoji Similarity Modeling to Support Expressive Communication. *ACM Transactions on Computer-Human Interaction* 24, 1, Article 6 (2017), 41 pages. https://doi.org/10.1145/3039685

[21] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine Mcleavey, and Ilya Sutskever. 2023. Robust Speech Recognition via Large-Scale Weak Supervision. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 28492–28518. https://proceedings.mlr.press/v202/radford23a.html

[22] Andreas Triantafyllopoulos, Anton Batliner, Simon Rampp, Manuel Milling, and Björn Schuller. 2024. INTERSPEECH 2009 Emotion Challenge Revisited: Benchmarking 15 Years of Progress in Speech Emotion Recognition. In *Interspeech 2024*. 1585–1589. https://doi.org/10.21437/Interspeech.2024-97

[23] Johannes Wagner, Andreas Triantafyllopoulos, Hagen Wierstorf, Maximilian Schmitt, Felix Burkhardt, Florian Eyben, and Björn W. Schuller. 2023. Dawn of the Transformer Era in Speech Emotion Recognition: Closing the Valence Gap. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 9 (2023), 10745–10759. https://doi.org/10.1109/TPAMI.2023.3263585

[24] Philip Weber, Lea Katharina Michel, Lena Koschorreck, and Thomas Ludwig. 2023. Voice Messages Reimagined: Exploring the Design Space of Current Voice Messaging Interfaces. In *Proceedings of Mensch Und Computer 2023* (Rapperswil, Switzerland) *(MuC '23)*. Association for Computing Machinery, New York, NY, USA, 336–340. https://doi.org/10.1145/3603555.3608562

[25] Sarah Wiseman and Sandy J. J. Gould. 2018. Repurposing Emoji for Personalised Communication: Why [Pizza Slice Emoji] Means "I Love You". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173726

[26] Chi-Lan Yang, Shigeo Yoshida, Hideaki Kuzuoka, Takuji Narumi, and Naomi Yamashita. 2023. Affective Profile Pictures: Exploring the Effects of Changing Facial Expressions in Profile Pictures on Text-Based Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 270, 17 pages. https://doi.org/10.1145/3544548.3581061