2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)

MoodFlow: Orchestrating Conversations with Emotionally Intelligent Avatars in Mixed Reality

Llogari Casas [*]
3FINERY LTD
Edinburgh Napier University

Samantha Hannah[†]
3FINERY LTD
Edinburgh Napier University

Kenny Mitchell[‡]
3FINERY LTD
Edinburgh Napier University

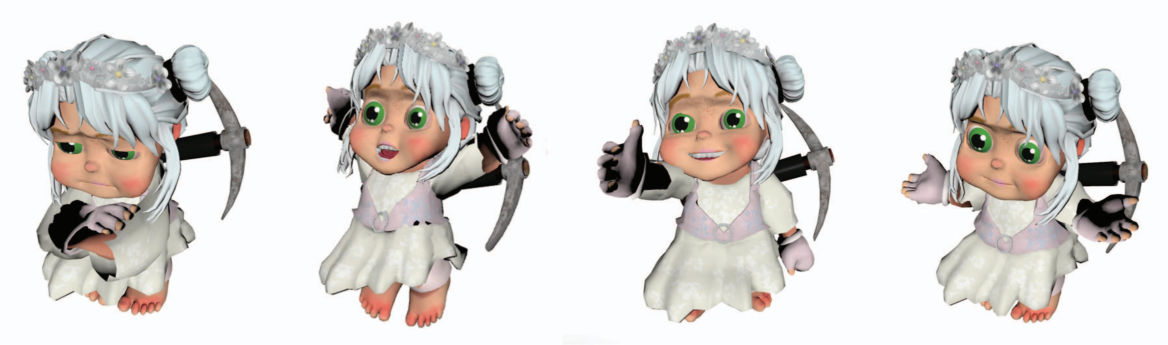

Figure 1: Kara, our avatar, exhibits a range of emotional reactions through dynamically triggered animations, responding to human conversations to convey emotions in a compelling and authentic manner.

Abstract

MoodFlow presents a novel approach at the intersection of mixed reality and conversational artificial intelligence for emotionally intelligent avatars. Through a state machine embedded in user prompts, the system decodes emotional nuances, enabling avatars to respond authentically to the spectrum of human emotions. Our system employs expressive avatars with a shared structure, allowing for seamless animation transferability between avatars with distinct outlook. The avatars, optimized for mixed reality, incorporate low-poly designs and toon shader stylization. This immersive journey transforms virtual conversations into open-ended dialogues, where avatars go beyond scripted interactions, adapting in real-time based on emotional context. Beyond entertainment, the approach envisions diverse applications, including virtual therapy, education, entertainment, corporate communication, and social interactions by opening doors to emotionally rich experiences across sectors.

Index Terms: H.1.2 [Models and Principles]: User/Machine Systems—Human-centered computing; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation; I.3.6 [Computer Graphics]: Methodology and Techniques—Methodology; K.8.0 [Personal Computing]: General—Games;

1 Introduction

In the ever-evolving landscape of Mixed Reality (MR), where the boundaries between physical and digital worlds blur, human-computer interactions stand at the forefront of innovation. As society traverses through an era saturated with digital interactions, the need for emotionally meaningful connections become increasingly pronounced. In response to this demand, *MoodFlow* embarks on a journey to redefine the dynamics of engagement amongst the reality–virtuality continuum [7]. At its core, *MoodFlow* is a convergence of state-of-the-art technologies, seamlessly blending conversational AI, mixed reality, and intelligent avatars. The goal is not merely to facilitate conversations within the immersive space of MR, but to elevate these interactions into a realm where emotions and dialogues seamlessly coexist.

The conventional boundaries of dialogue are often restricted by the limitations of existing virtual communication tools, resulting on traditional avatars lacking the nuanced understanding required to convey the intricacies of human emotions. In this immersive experience, each interaction becomes a open-ended conversation guided by emotionally intelligent avatars that respond dynamically to human expressions. *MoodFlow* unveils a novel approach by embedding a state machine within user prompts. This mechanism becomes the bridge between users and avatars, translating emotional cues into a language that avatars can comprehend. The result is an immersive journey where every word carries not just meaning, but a spectrum of emotions associated to it.

2 Emotionally Intelligent Avatars

Traditional avatars, often constrained by predefined animations and limited responsiveness, often fall short in capturing the intricate nuances of emotional expression. This limitation not only hinders the depth of virtual interactions, but also diminishes the potential for meaningful connections in digital spaces. Humans are inherently emotional beings, and a significant portion of our communication relies on non-verbal cues, facial expressions, and body language. In virtual environments, where physical presence is absent, the absence of genuine emotional expression in avatars has been a persistent obstacle. Users often find it challenging to connect on a deeper level and feel fully immersed in virtual experiences when their avatars lack the ability to express emotions dynamically.

Emotionally Intelligent Avatars (EIAs) emerge as a solution to

[*] e-mail: llogari@3finery.com
[†] e-mail: samantha@3finery.com
[‡] e-mail: kenny@3finery.com

DOI 10.1109/VRW62533.2024.00021

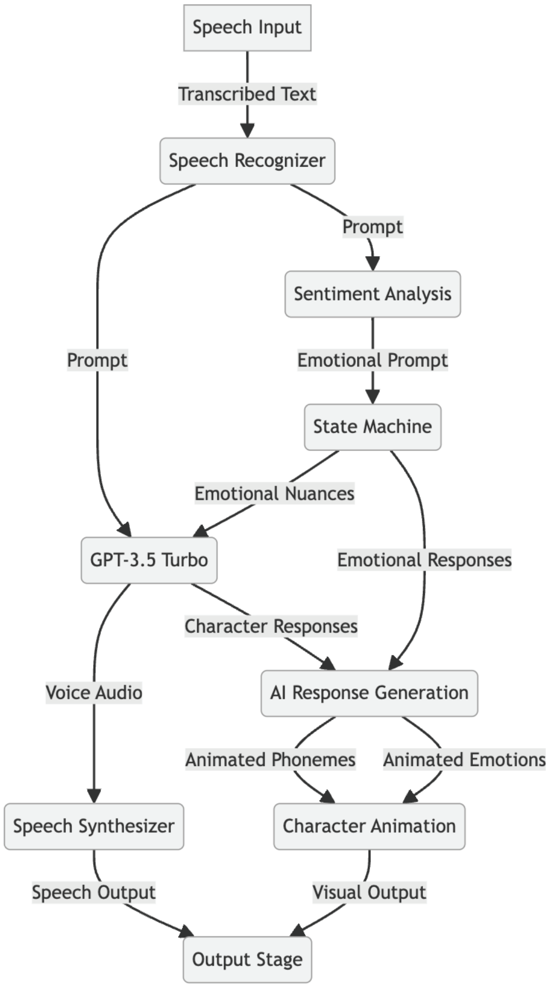

Figure 2: Flowchart depicting the interaction and data flow in MoodFlow for real-time speech processing and avatar animation. Speech input is transcribed, prompting both AI-driven character responses and emotional analysis. The resulting animated phonemes and emotions are synthesized into speech and visual output for an immersive user experience.

infuse virtual interactions with the authenticity and depth that have been lacking thus far. By enabling avatars to interpret and respond to users' emotional states, we can enhance the quality of virtual conversations, making them more meaningful, expressive, and conducive to genuine human connection. Recent years have witnessed significant advances in Natural Language Processing (NLP), contributing to the development of EIAs. Recent works by Mendes et al. [8], have explored the application of advanced sentiment analysis algorithms to enhance the avatar's ability to comprehend and respond to user emotions expressed in textual inputs. The integration of computer vision technologies for real-time facial expression analysis is a key area of focus in EIAs. Pioneering efforts have demonstrated the effectiveness of high-quality facial animation and expression recognition in creating emotionally expressive avatars [5]. Previous research has emphasized the importance of adapting their behaviour based on user emotions and contextual cues [4], making interactions more personalized and responsive. The infusion of empathy and social skills into avatars has also been a focus of research, as evidenced by works assessing its effectiveness [6]. A growing area of application for emotionally intelligent avatars lies in mental health and therapy. Recent contributions explore the potential of avatars to provide non-judgmental interactions and support in therapeutic interventions [9]. Building upon our previous work, in which we introduced AI conversational characters in *Intermediated Reality* [2], *MoodFlow* integrates a state machine into the responses generated by a LLM to decode subtle emotional nuances embedded within user prompts and provide an emotionally intelligent conversational artificial agent.

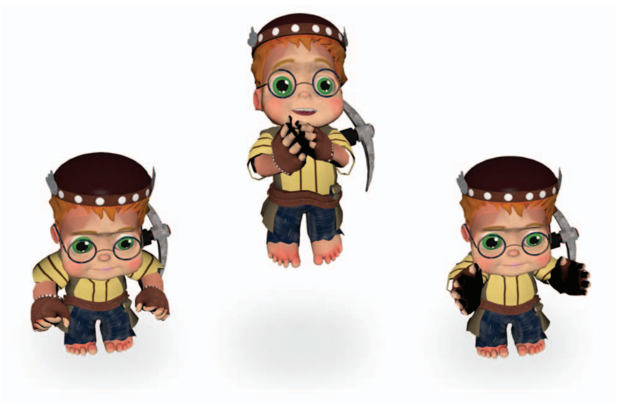

Figure 3: *Quin* expressing emotional joy through a key-frame based animation, showcasing the efficiency and effectiveness of the technique by only utilizing three key frames.

3 MoodFlow

With *MoodFlow*, as participants engage in discussions, the avatars actively respond to the collective emotional atmosphere, providing a visual representation of the sentiments within the virtual space as depicted in figure 1. This not only enhances the authenticity of interactions but also facilitates a deeper level of empathy and connection. The avatars act as emotional mirrors, allowing participants to gauge the emotional tone of the conversation and fostering a more immersive and engaging communication experience.

Our current working prototype is implemented using *Swift* programming language. As illustrated in Figure 2, the *MoodFlow* system begins with the user speaking into the microphone and capturing the spoken prompt. This input is then transcribed into text using the *SFSpeechRecognizer* library [1]. This text subsequently serves as a prompt for two distinct processes: the generation of character responses through the Large Language Model API (GPT-3, 3.5-turbo [3]) for cloud processing and the analysis of emotional content via sentiment analysis using *NLTagger* library [1]. The state machine is implemented as an array, sorting emotional states in order of confidence as outputted by the sentiment analyzer. This array is reset for each prompt, ensuring that only the emotional state of the current prompt is analyzed. The state machine dynamically processes these emotional nuances, guiding the generation of character responses to align with the prevailing emotional context by sourcing that information to the LLM. Following that, the AI dynamically produces brief character responses, which subsequently undergo processing to transform them into animated phonemes [10]. Each phoneme is smoothly integrated with their immediate and previous tokens, allowing for a natural blending of visemes while controlling its blend weights acceleration to ensure physical plausibility. These visemes are then utilized to animate the facial rig parameters of the characters, ensuring precise synchronization with the generation of voice audio. The resulting AI-generated response is sourced to the *AVSpeechSynthesizer* library [1] for speech generation, concur-

Figure 4: *Kara*, our emotionally intelligent avatar, warmly greets the user in a Virtual Reality setting. The immersive experience begins with *Kara*'s friendly welcome, setting the tone for a dynamic and emotionally rich interaction.

rent with procedurally animating the avatar's emotional reactions based on the current status of the AI-generated response and the state-machine [1]. The speech recognition and synthesis libraries operate on the device using hardware-accelerated methods, ensuring real-time responsiveness, while the LLM leverages cloud resources for processing responses.

4 Animated Emotions

Our avatars are purposefully designed for MR scenarios, featuring a low-polygonal mesh for optimal real-time playback. The avatars employed in the experience, such as *Kara* or *Quin*, underwent a dual-phase creation process. Initially sculpted using traditional 3D modelling software, these came to life through a series of concise animated clips, each depicting a distinct emotional response. These animations intricately correspond to different states within the embedded state-machine, representing various emotional phases. The resulting animated emotions are a fusion of key-framed animations and deliberate character stylizations. Inspired by the desire to create emotionally-driven avatars, our design approach focuses on conveying emotions through both appearance and movement. One of the distinctive features lies in the intentional reduction of key frames as depicted in figure 3. This approach serves a dual purpose: it enhances the characters' expressiveness while significantly streamlining the animation development process. By employing fewer key frames, we maintain efficiency in creating a diverse range of emotional responses and we give a unique and representative visual appeal to the experience.

To enhance visual appeal, we employ a toon shader effect, providing a stylized, cartoon-like appearance. A key aspect of our avatar design is the shared structure between both the girl and boy avatars. This extends beyond a common visual framework to encompass identical body structures, including rigging and the allocation of vertices. This deliberate uniformity brings the advantage to allow transferability of animations between avatars with different outlook as depicted in figure 5. With the same underlying body structure, animations seamlessly transition between different customization combinations. All animations within our framework maintain a consistent duration, ranging between 2 and 5 seconds. Each animation shares a common rest pose facilitating seamless blending between animations. This shared starting point enables a smooth transition between different emotional states, ensuring that the avatar's movements flow naturally and authentically. The system achieves a seamless emotion continuum by queuing the emotional responses in the animation engine following a sequential scheduling approach.

Utilizing the common body structure among avatar geometries, both clothing and body textures share identical texture coordinates across variations. This streamlined approach simplifies the integration of new outfits, requiring only the addition of a new texture map adhering to established coordinates to introduce additional combinations. As depicted in figure 6, the outcome is a wide range of diverse appearances easily integrated into the experience, providing users with an extensive array of choices without compromising the overall design coherence.

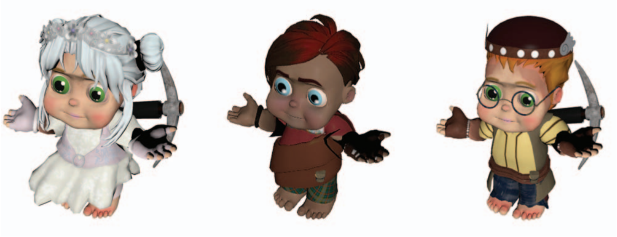

Figure 5: *Kara*, *Niamh* and *Quin* express emotional hopelessness using the same animation, leveraging the shared body structure. This enables the seamless transfer of animations between avatars with distinct outlooks.

5 Applications

The versatile capabilities of *MoodFlow* extend beyond entertainment, offering potential applications in various domains. In the realm of virtual therapy, emotionally intelligent avatars could create an immersive environment for therapeutic conversations. The nuanced understanding of emotions would provide a more empathetic and personalized experience for users. Similarly, in an educational setting, this approach has the potential to revolutionize student interactions by offering dynamic and engaging explanations, adapting to students' emotional states. This not only would enhance the learning experience but also create a more interactive and responsive educational environment. Further, the entertainment industry, including gaming, stands to benefit from the emotionally rich experiences offered by this approach. Characters with authentic emotional responses would add depth to gaming narratives, making virtual worlds more immersive and captivating. In the corporate landscape, *MoodFlow* could find applications in virtual meetings and training sessions. This feature would not only enrich the expressive nature of the conversation, but would also add a layer of emotional intelligence to multi-person remote interactions, making the communication more engaging, relatable, and immersive.

6 Discussion & Conclusions

In our working prototype, the incorporation of a state machine has empowered our avatars to authentically respond to a diverse range of human emotions, overcoming the traditional limitations of scripted interactions. As observed in preliminary feedback from users, one of the key advantages lies in the ability of the system to bridge the emotional gap created by physical separation, as often experienced in traditional videoconferencing tools. The deliberate design choices, such as a shared structure for avatars, low-poly optimization, and toon shader stylization, contribute to an engaging user framework that converges more towards a social-game experience, than to a traditional conversational platform.

While *MoodFlow* represents a novel approach in the realm of emotionally intelligent avatars, it is important to acknowledge certain limitations. The system's effectiveness heavily relies on the accuracy of emotional cues decoded from user prompts, and there may be instances where misinterpretations occur. Additionally, the current version of *MoodFlow* focuses primarily on a predefined set of emotions, and expanding this repertoire to encompass a broader range remains a potential avenue for improvement. Furthermore, the

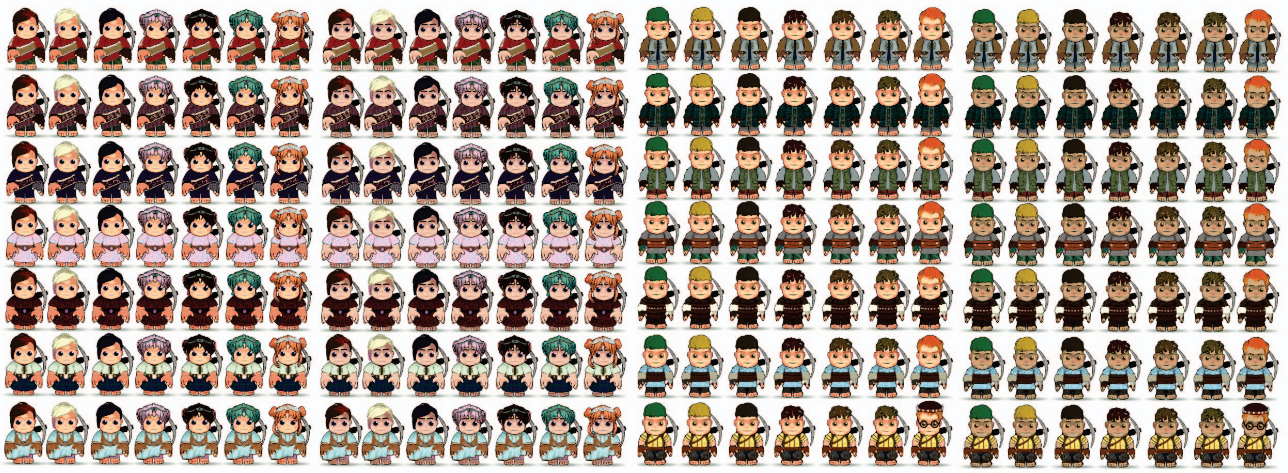

Figure 6: A visual representation highlighting the versatility of *MoodFlow*'s customization feature, showcasing a diverse array of avatar appearances achieved through unique combinations of clothing, hairstyles, and accessories.

system's performance may vary based on individual user characteristics, such as accent or speech patterns, influencing the accuracy of emotional analysis. Expanding the emotional repertoire could involve incorporating more nuanced emotional states and leveraging user feedback to refine the system's responses. Additionally, considering individual user characteristics in the training data and employing accent-agnostic models could contribute to a more inclusive and accurate emotional analysis. Additionally, advancements in artificial intelligence and mixed reality technologies could further enhance the depth and complexity of emotional interactions, making the avatars even more responsive and intuitive. The user customization aspect could be expanded to include more diverse options, providing users with an even broader range of choices to personalize their avatars that could promote further facial, body and accessories diversity.

In our roadmap for future work, we envision a comprehensive enhancement of the capabilities and applicability of the *MoodFlow* system, extending its impact to real-world scenarios. This will involve a thorough exploration of some of the applications described in section 5 to evaluate its performance and effectiveness in diverse settings. By subjecting the system to real-world contexts, we aim to gather valuable empirical data that could inform improvements and adaptations. Concurrently, user studies will play a crucial role in this process, providing deeper insights into user perceptions, preferences, and interactions with the system. This multifaceted approach, combining practical applications and user studies, will contribute not only to the refinement of the existing system, but also to the development of guidelines for its optimal usage across different domains. Through this iterative and user-centric methodology, we anticipate the evolution of *MoodFlow* into a versatile and robust platform that could address a spectrum of real-world communication challenges in the future.

Acknowledgments

The authors would like to express their gratitude to Rachel Robinson, Sam Flint and Alex Cullinane-Carrasco for their contributions in conceptualizing, developing and animating the avatars used in this research article. This project has received funding from InnovateUK's grant agreement number 68475 and from European Union's Horizon 2020 research and innovation programme under grant agreement number 101017779.

References

[1] Apple. https://developer.apple.com/documentation/, 2024.

[2] L. Casas and K. Mitchell. Intermediated reality with an ai 3d printed character. In *ACM SIGGRAPH 2023 Real-Time Live!*, SIGGRAPH '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3588430.3597251

[3] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, W. Ye, Y. Zhang, Y. Chang, P. S. Yu, Q. Yang, and X. Xie. A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.*, jan 2024. Just Accepted. doi: 10.1145/3641289

[4] M. Fabri, D. J. Moore, and D. J. Hobbs. The emotional avatar: Nonverbal communication between inhabitants of collaborative virtual environments. In *International gesture workshop*, pp. 269–273. Springer, 1999. doi: 10.1007/3-540-46616-9_24

[5] M. Gonzalez-Franco, A. Steed, S. Hoogendyk, and E. Ofek. Using facial animation to increase the enfacement illusion and avatar self-identification. *IEEE transactions on visualization and computer graphics*, 26(5):2023–2029, 2020. doi: 10.1109/tvcg.2020.2973075

[6] E. Johnson, R. Hervás, C. Gutiérrez López de la Franca, T. Mondéjar, S. F. Ochoa, and J. Favela. Assessing empathy and managing emotions through interactions with an affective avatar. *Health informatics journal*, 24(2):182–193, 2018. doi: 10.1177/1460458216661864

[7] S. Mann. Mediated reality. *Linux J.*, 1999(59es):5–es, mar 1999. doi: 10.5555/327697.327702

[8] C. Mendes, R. Pereira, J. Ribeiro, N. Rodrigues, and A. Pereira. Chatto: An emotionally intelligent avatar for elderly care in ambient assisted living. In *International Symposium on Ambient Intelligence*, pp. 93–102. Springer, 2023. doi: 10.1007/978-3-031-43461-7_10

[9] E. Miller and D. Polson. Apps, avatars, and robots: The future of mental healthcare. *Issues in mental health nursing*, 40(3):208–214, 2019. doi: 10.1080/01612840.2018.1524535

[10] M. S. Yavas. *Phonetics*, chap. 1, pp. 1–29. John Wiley Sons Ltd, New York, NY, USA, 2011. doi: 10.1002/9781444392623.ch1