

# SentimentVoice: Integrating Emotion AI and VR in Performing Arts

Semi Ryu  
Kinetic Imaging  
VCU Arts  
Richmond, VA, USA  
sryu2@vcu.edu

Miles Popiela  
Computer Science  
VCU Engineering  
Richmond, VA, USA  
popielamc@vcu.edu

Henry Bryant  
VCU Arts  
Richmond, VA, USA  
hcbryant9@gmail.com

Matthew Labella  
Kinetic Imaging  
VCU Arts  
Richmond, VA, USA  
Matthewlabellar@gmail.com

Hanna Chou  
Kinetic Imaging  
VCU Arts  
Richmond, VA, USA  
hannachou02@gmail.com

Christine Rayburn  
Kinetic Imaging  
VCU Arts  
Richmond, VA, USA  
rayburnc@vcu.edu

Alberto Cano  
Computer Science  
Virginia Tech  
Blacksburg, VA, USA  
acano@vt.edu

## Abstract

SentimentVoice is a live VR performance art project that repurposes emotion-tracking AI to amplify immigrant narratives. While emotion AI is often used for surveillance or commercial gain, SentimentVoice subverts this technology into an empathetic tool, actively listening to immigrant stories instead of exploiting them. The project integrates human voice and facial expression as core materials, analysed through emotion AI using face tracking, the ChatGPT API, voice analysis, and speech-to-text conversion. While technologies like lie detectors and facial recognition have been used to monitor and target marginalized communities, SentimentVoice reclaims emotion data as a medium for fostering empathy, deep connections, and mutual understanding. Rooted in oral history, the project collects real immigrant stories from the Richmond community, structured around five key locational memories: airports, schools, immigration offices, hospitals and workplaces. These spaces represent critical moments of transition and adaptation in the immigrant experience.

## CCS Concepts

• **Human-centered computing, Affective computing;**

## Keywords

VR, Emotions, Sentimental analysis, Oral history, AI

## ACM Reference Format:

Semi Ryu, Miles Popiela, Henry Bryant, Matthew Labella, Hanna Chou, Christine Rayburn, and Alberto Cano. 2025. SentimentVoice: Integrating Emotion AI and VR in Performing Arts. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Immersive Pavilion (SIGGRAPH Immersive Pavilion '25)*, August 10–14, 2025, Vancouver, BC, Canada. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3721245.3734053>

## 1 PROJECT DESCRIPTION

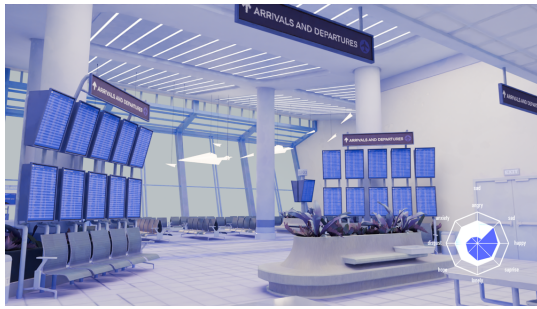
The SentimentVoice is a live VR performance art project that utilizes emotion-tracking AI technology to amplify the narratives of immigrant populations. Inspired by the pervasive use of emotional surveillance and the vulnerabilities surrounding emotion data, SentimentVoice subverts emotion AI technology, typically used for surveillance or commercial purposes, into an empathetic mediator for actively listening to the stories of immigrants. The project incorporates human voice and movement as the core material for the performance art, which is detected and analyzed by emotion AI. This analysis leverages various technologies, including face tracking, the ChatGPT API, voice analysis, and speech-to-text conversion. Emotion-tracking AI, such as lie detectors and facial recognition, has often been used to create monitored environments, frequently targeting minority populations. However, it is important to recognize that emotion data can also be harnessed to cultivate empathy, foster deep connections, and promote mutual understanding.

SentimentVoice began with an oral history process, gathering authentic stories from many immigrants in the Richmond community, based on five locational memories: airports, schools, immigration offices, hospitals, and workplaces.

During the live performance, one actor wears a VR headset while the other does not, establishing an intricate connection between virtual and actual space, as well as between past memories and present life. As the actors navigate the five VR locations and tell the associated stories, Emotion AI responds to their speech with

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
*SIGGRAPH Immersive Pavilion '25, Vancouver, BC, Canada*  
© 2025 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-1547-1/2025/08  
<https://doi.org/10.1145/3721245.3734053>



**Figure 1: VR environment with airport visuals. Render by Matthew LaBella.**



**Figure 2: VR environment of a school hallway. Render by Matthew LaBella.**

visual (particles, textures, lights, etc.) and sound acknowledgement within the VR environment. The process of emotion tracking is displayed to demonstrate how our speech and facial expressions are being detected without our awareness, highlighting cybersecurity issues.

## 2 Technology

This project focuses on a multi-modal emotion recognition system specifically designed for VR environments. A key component is a multi-threaded audio processing system developed in Unity. This system continuously captures microphone data from multiple devices, ensuring real-time responsiveness and optimized performance. The captured audio is then processed by Whisper for speech-to-text transcription. This transcribed dialogue is dynamically saved and analysed by a Large Language Model (LLM), specifically ChatGPT-4o mini, to quantify a range of emotions (Angry, Sad, Happy, Surprise, Lonely, Hope, Disgust, and Anxiety) on a scale of 0-10. In parallel, the system incorporates a facial emotion detection model. This model, a trained random forest classifier, utilizes the Facial Action Coding System (FACS)-based blendshape output values from the Face Tracking for Movement SDK for Unity as its input. Notably, this model was trained on a self-procured dataset due to the scarcity of publicly available datasets for this specific VR use case. The emotion values derived from speech are then compared with the outputs from this facial emotion detection model.

The system then determines the predominant emotion, which dictates the activation and intensity of corresponding visual effects within the VR environment. If two emotions are detected with equal intensity, their respective visual activations are combined, allowing for the simultaneous display of multiple emotional states. This approach allows for a nuanced understanding of emotional states by integrating and comparing data from both vocal and facial expressions and translating these detected emotions into dynamic visual feedback within a VR setting.

## 3 Full Credits

Creative Director: Semi Ryu

AI Director: Alberto Cano

AI Engineer/VR Developer: Miles Popiela, Henry Bryant

3D Modeler and Texture Artist: Matthew LaBella

Video Editor: Hanna Chou

Sound Artist: Chrystine Rayburn

Actors: Ryan Flores, Katherine Nguyen

Script Writer: Hanna Chou, Katherine Nguyen, Ryan Flores /

Videographer: Hanna Chou, Kiara Brown, Uday Illa, Ryan Alvarado

/ Interviewer: Hanna Chou, Richmond Animation Archive, Menna

Hassanain/ Assistant Programmer: Ariana Thomas and Josiah Wil-

son

## Acknowledgments

We would like to thank the Commonwealth Cyber Initiative for their support in the development of this project. Special thanks to everyone who generously shared their stories