UIST 2025 Poster

PerformaVis: Real-Time Affective Music Visualization Driven by Pianist's Bodily Expressions

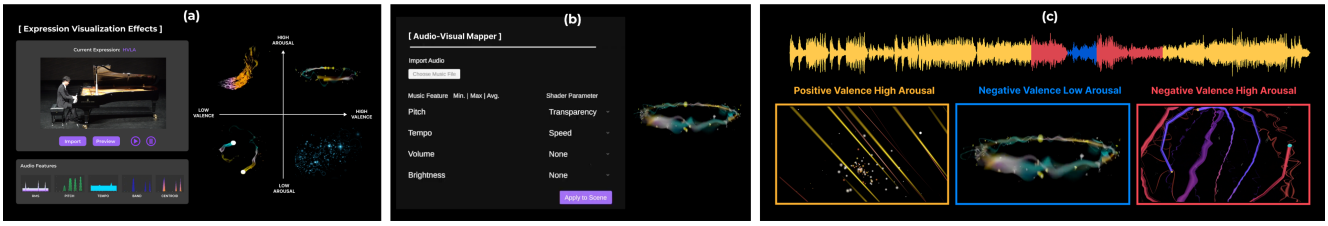
Tai-Chen Tsai[1], Chih-Chuan Huang[1], Shun-Han Chang[1], Cheng-Yin Hsu[1], Hsin-Ying Lee[1], Kai-Hsiang Wen[2], Tse-Yu Pan[3] and Min-Chun Hu[1]

[1] National Tsing Hua University, Taiwan [2] Glisten Arts, Taiwan [3] National Taiwan University of Science and Technology, Taiwan

INTRODUCTION

Current music visualization systems focus primarily on audio features, overlooking the affective interpretations conveyed by performers' bodily expressions during live performance. Different performers interpret the same piece with varying affective nuances, yet existing systems cannot capture these individualized expressions. We introduce PerformaVis, a real-time system that recognizes pianists' affective states through computer vision, categorizes them using the affect circumplex model, and integrates them with audio analysis. Our multi-layered editor enables customizable affect-to-visual mappings, helping audiences perceive both musical and affective nuances in live performance.

METHOD



Figrue1. PerformaVis workflow. (a) Input processing with emotion recognition and visual mapping. (b) User-defined audio-visual parameter mappings. (c) Real-time visual effects driven by affective states and audio features

PerformaVis consists of three main modules:

- **Emotion Recognition Module**: We constructed a specialized 156-minute multimodal dataset from 5 experienced pianists, capturing synchronized audio, RGB-D video from top and side views, and wrist-worn IMU data with self-annotated affective states. Our TSM-based RGB model achieves optimal performance with 1.11-second inference time, enabling real-time detection of affective states from performers' bodily expressions during live performance.
- **Audio Feature Extraction Module**: We developed an audio analysis system that extracts key features from both time and frequency domains. Time domain features include tempo for rhythmic pacing and RMS energy for loudness variations. Frequency domain features capture pitch for fundamental frequency and spectral centroid for timbral brightness. These features provide the foundation for personalized audio-visual mappings that respond to the acoustic qualities of the music.
- **Multi-layered Visualization Editor**: Our visualization editor features two configuration layers: the first maps visual effects to affective quadrants based on the circumplex model as shown in Figure 1(a), while the second allows custom relationships between audio features and visual parameters as shown in Figure 1(b). The system includes a default mode where tempo controls timing, RMS energy controls intensity, pitch modulates hue, and spectral centroid adjusts brightness. This enables real-time visual synchronization with both performer affective expression and musical characteristics as shown in Figure 1(c).

PILOT STUDY

- **Participants:** We tested our system with 14 participants (8M, 6F; mean age = 24.4 years) from diverse academic backgrounds.
- **Experimental Design:** Participants experienced three conditions in randomized order: Audio Only, Audio Visualization (audio features), and Affective Audio Visualization (our system), then completed post-experience questionnaires using 5-point Likert scales.
- **Evaluation Method:** The questionnaire included the NASA Task Load Index for cognitive workload measurement and eight custom dimensions for user experience assessment.
- **Key Findings**: While audio-only achieved highest focus scores, both visualization methods performed similarly in meaningfulness and understanding. However, our affective approach specifically excelled in affective engagement (4.21±0.45) and interest enhancement (3.71±0.9), demonstrating that integrating performer emotions creates more engaging musical experiences than traditional audio-driven visualizations.

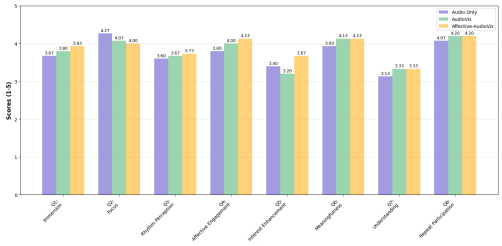


Figure 2. Pilot study results comparing three conditions across eight evaluation dimensions.

DISCUSSION & FUTURE WORK

Our results demonstrate that performer-driven affective cues significantly enhance audience engagement more than audio-driven visualizations Current limitations include scarce annotated affective movement data. Future work will explore multimodal fusion to improve emotion recognition robustness and accuracy.