The Tenth International Conference on Neuroscience and Cognitive Brain Information
BRAININFO 2025
AAHED: Analysis and Applications of Human Emotion Dynamics

Dynamic Emotion Analysis in Piano Music Based on Performance Techniques Recognition

Authors: Yena Wu, Clement Leung

School of Science and Engineering

the Chinese University of Hong Kong, Shenzhen

121090620@link.cuhk.edu.cn, clementleung@cuhk.edu.cn

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Yena WU

School of Science and Engineering

the Chinese University of Hong Kong, Shenzhen

Yena WU is currently a senior student majoring in Computer Engineering at The Chinese University of Hong Kong, Shenzhen.

Her research interests lie in the fields of artificial intelligence, algorithms, and related areas.

Background and Motivation

- **Music and Emotion:**
  Music as a "language of emotions"; significance in musicology and psychology.
- **Limitations of Existing Work:**
  - Focus on lyrics, dynamics, and volume, ignoring the impact of performance techniques.
  - Holistic emotion assessment overlooks dynamic emotional changes in music segments.
- **Research Objectives:**
  - Automatically identify piano performance techniques using deep learning.
  - Analyze how these techniques influence dynamic emotional expression.

Methodology Overview

A. Data Preprocessing

raw audio → format conversion → segmentation → augmentation

Format Standardization:
- Convert all audio files to WAV format.
- Resample to 44.1 kHz, 16-bit depth, mono channel.

Audio Segmentation: Divide audio into 1.5s and 3s segments.
- 1.5s segments: For rapid techniques (e.g., glissando, octave).
- 3s segments: For longer techniques (e.g., arpeggio, vibrato).

Data Augmentation:
- Time Shifting: Randomly shift audio by $\pm$ 500 samples to simulate variations.
- Gaussian Noise: Add noise (σ = 0.5% of signal amplitude) to improve robustness.

Methodology Overview

B. Feature Extraction and Stacking

1.Glissando Features:
- Delta (1st-order derivative) and Delta-Delta (2nd-order derivative) features.

2. Vibrato Features:
- Modulation frequency and amplitude features.

3. Octave Features:
- Harmonic analysis to capture frequency relationships.

4. Arpeggio Features:
- Delta, Delta-Delta, and time interval features.

Combine features into multi-channel tensors (e.g., Mel-spectrogram + Delta + Delta-Delta for glissando).

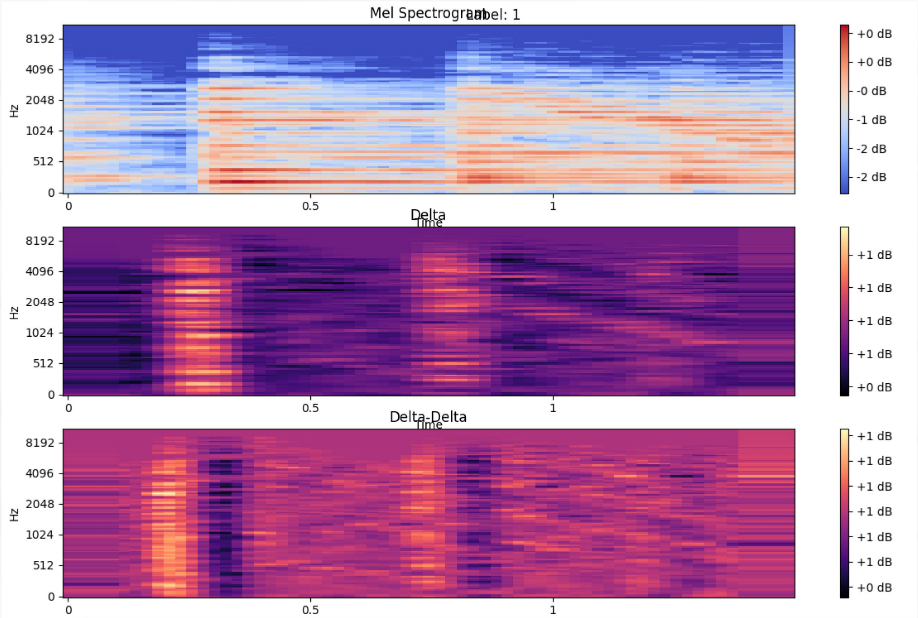


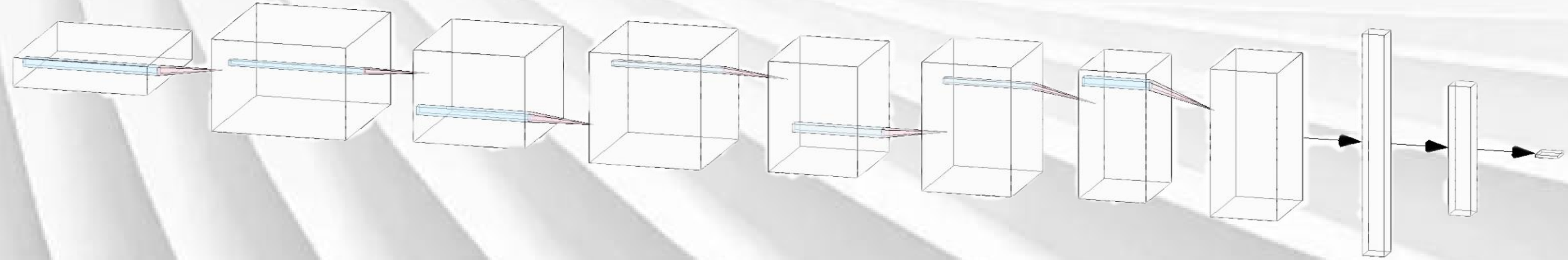
Figure 1. glissando features

Methodology Overview

C. Performance Technique Recognition Model

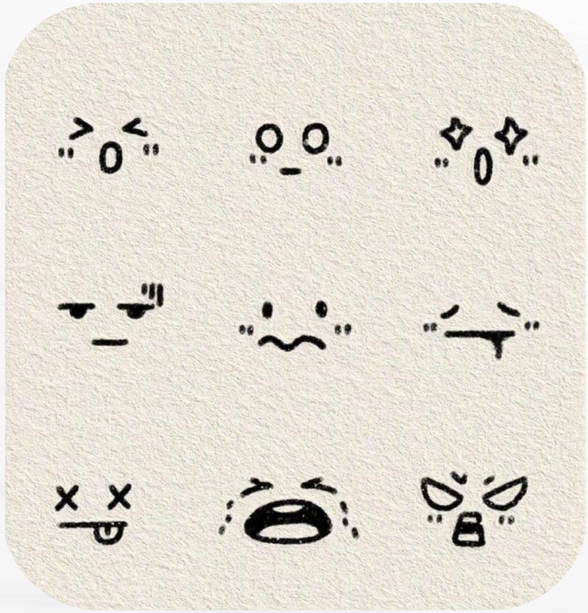
CNN Architecture

- Input Layer:
  - 3-channel Mel-spectrogram (128 Mel bands, 65 frames).
- Convolutional Layers:
  - 4 layers with 3x3 kernels, followed by max-pooling (2x2).
- Adaptive Pooling:
  - Adaptive average pooling to reduce feature maps to 1x1 size.
- Fully Connected Layers:
  - 2 layers with ReLU activation and dropout for regularization.
- Output layer with sigmoid activation for binary classification.



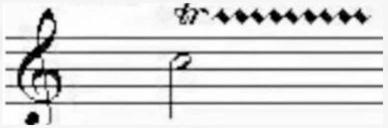
Methodology Overview

E. Emotion Analysis

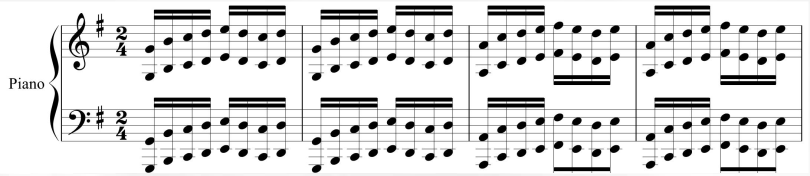


- **Emotion Labeling:**
  - Use GEMS (Geneva Emotional Music Scales) with 45 emotion tags across 9 categories.

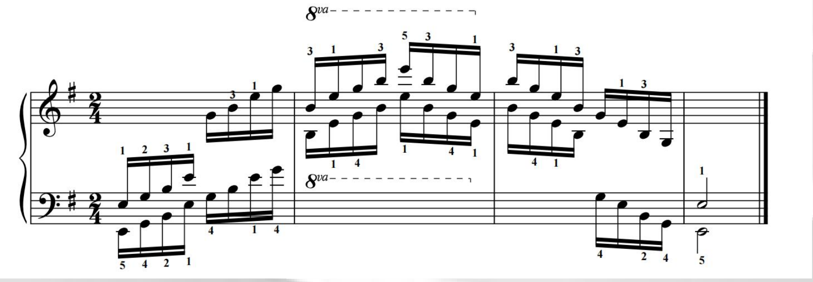
Amazement, Solemnity, Tenderness, Nostalgia, Calmness, Power, Joyful Activation, Tension, and Sadness.



- **Pearson Correlation Analysis:**
  - Quantify relationships between performance techniques and emotions.
  - Example: Vibrato ➔ "Tension" (positive correlation).



- **Dynamic Emotion Analysis:**
  - Track emotional changes over time in 3s segments.
  - Weight emotions by decibel levels of techniques.



Result

A. Performance Metrics

TABLE II
PERFORMANCE METRICS FOR PIANO PERFORMANCE TECHNIQUES

Technique	Accuracy	Precision	Recall	F1-score
Glissando	89.5%	88.3%	87.6%	89.9%
Octave	86.2%	88.1%	84.7%	86.4%
Arpeggio	83.0%	82.9%	84.3%	83.1%
Vibrato	85.8%	83.7%	88.2%	85.9%

Glissando performed best in accuracy, accuracy, recall and F1 score, especially in the F1 score of 89.9%. Octave accuracy is the highest at 88.1%, but the overall F1 score is slightly lower than that of the glissando. Arpeggios performed the worst among the indicators, with the lowest accuracy of 83.0%. The vibrato performed better in recall and F1 scores, but still fell short of the glissando and octaves. Overall, glissand is the most recognizable technique.

Result

B. Pearson Correlation Analysis Results

TABLE III

PEARSON CORRELATION COEFFICIENTS BETWEEN PERFORMANCE TECHNIQUES AND EMOTIONS.

Performance Technique	Joyful Activation	Calmness	Tension	Amazement	Sadness	Solemnity	Power	Tenderness	Nostalgia
Glissando	0.65	0.21	0.71	0.55	-0.31	-0.47	0.62	-0.20	-0.60
vibrato	0.30	-0.25	0.65	0.53	0.65	-0.40	-0.35	0.78	0.80
Arpeggio	0.62	-0.10	-0.26	-0.55	-0.32	-0.30	0.13	0.73	0.82
Octave	0.75	-0.45	0.60	0.60	-0.50	0.54	0.90	-0.80	-0.56

Glissando has a strong positive correlation with pleasure activation, surprise and power. Vibrato are highly associated with nostalgia and tenderness, and are positively associated with sadness. Arpeggios were positively correlated with nostalgia and tenderness, but negatively correlated with tension and sadness. The octave shows a strong sense of power and pleasure activation, and is negatively associated with tenderness and sadness.

Result

C. Dynamic Emotion Analysis Results

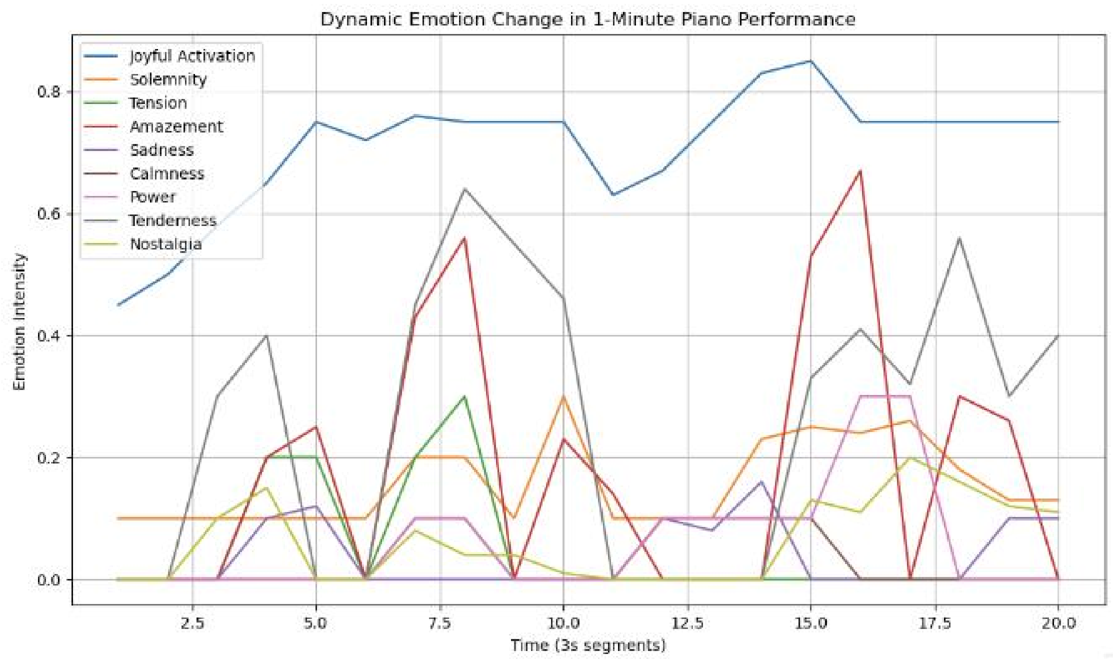


Figure 2. Dynamic Emotion Change in 1-Minute Piano Performance

Results of the dynamic emotion analysis applied to the performance of Czerny Op. 365 No.33, a Polish dance：

- The octave technique, present in the majority of the segments, was predominantly associated with the emotion of joyful.
- The vibrato technique, observed in segments 4, 5, 13, 15, 16, 17, 19, and 20, was associated with emotional expressions such as amazement, tension, and sadness.
- The glissando technique, detected in segments 7 and 10, elicited emotions like joyful and activation.

Discussion

Success points:

- High-precision skill recognition provides reliable input for dynamic sentiment analysis.
- The strong correlation between emotion and skill validates the theoretical hypothesis of musical expression.

Two limitation：

- Overlapping Spectral Features: Techniques with similar harmonic patterns, such as arpeggios and trills, are occasionally misclassified. For instance, trills involve rapid note alternations that may overlap with arpeggio harmonics in the Mel-spectrogram.
- Independent Technique Detection: The current framework processes each technique independently, leading to redundant computations. A unified multi-label classification approach could better capture inter-technique dependencies (e.g., vibrato often co-occurs with legato phrasing).