Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 14, No. 1, 2025

Pipa Performance Analysis and Timbre Simulation Based on Artificial Intelligence

Yuqing Chen

Graduate University of Mongolia, Ulaanbaatar, 16052, Mongolia

Abstract: As a traditional Chinese musical instrument, pipa is famous for its unique timbre and complex and varied playing techniques. It is an important symbol of Chinese music culture. However, its complexity also makes its digital protection and research face many challenges. There are many kinds of playing techniques for pipa, such as finger rolling, plucking, bending, sweeping and overtones, each technique gives it a unique timbre change. In addition, the timbre of the pipa presents a variety of expressions due to the playing force, speed and the performer's skills, which puts higher demands on traditional music research and simulation technology.With the rapid development of artificial intelligence technology, especially breakthroughs in audio signal processing and deep learning, new solutions have been provided for the research of pipa art. By combining machine learning models and generative adversarial networks (GAN), significant progress can be made in technique analysis and timbre simulation. This study aims to use artificial intelligence technology to analyze the performance characteristics of the pipa and generate realistic timbre simulations, providing technical support for the digital protection and dissemination of traditional music.

Keywords: Analysis of pipa techniques, Tone simulation, Artificial intelligence applications.

1. Research Methods

1.1. Data collection and preprocessing

In order to achieve the classification and timbre simulation of pipa playing techniques, this study conducted comprehensive large-scale data collection. By using high-sensitivity condenser microphones, high-fidelity pipa playing audio was recorded, covering a variety of common playing techniques such as vibrato, plucking, and glissando. At the same time, an advanced motion capture system was used to record the performer's hand movement data with high precision, including fingering, force changes, and string position selection. These complementary datasets ensure the comprehensiveness of the acoustic and kinematic information required for subsequent analysis.

The audio data was carefully manually annotated by experts, recording the corresponding playing techniques, pitch, and rhythm information, providing reliable ground truth data for supervised learning tasks. In addition, the Librosa library [1] was used to extract detailed audio features, including spectrograms, Mel-frequency cepstral coefficients (MFCC), chromaticity features, pitch dynamics, and zero-crossing rate. These features, as input representations of machine learning models, fully capture the spectral and temporal characteristics of audio signals, providing a rich information basis for classification and simulation tasks.

1.2. Technique Classification Model

In order to automatically classify pipa playing techniques, this study designed a hybrid model based on convolutional neural network (CNN) [2] and recurrent neural network (RNN). CNN is used to extract spatial features of audio spectrograms, including harmonic structure, transient events and frequency distribution, while RNN captures the temporal series dynamics during the performance, such as note continuity and rhythm changes. This hybrid model that combines spatial and temporal features can effectively improve the recognition ability of complex playing techniques.

The training dataset of the model is divided into 70% for training, 20% for validation, and 10% for testing to ensure the reliability of the results. During the model training process, the cross entropy loss function is used to optimize the classification performance, and the learning rate scheduling is introduced to dynamically adjust the model learning progress. In addition, the early stopping strategy is used to prevent overfitting, and the robustness of the model is further improved by combining data enhancement methods such as time stretching, pitch shifting and noise injection of audio data.

The classification ability of the model is verified by confusion matrix analysis, especially in the classification of fingering and harmonics, but there is still room for improvement in the distinction between pitch bending and plucking. Future improvements include using attention mechanisms to enhance the model's sensitivity to complex techniques, and further improving classification accuracy through multimodal input (such as combining motion capture data).

1.3. Tone Simulation Model

In terms of timbre simulation, this study used a generative adversarial network (GAN) [3]. The GAN architecture consists of a generator and a discriminator, where the generator is responsible for generating a spectrogram that simulates the pipa timbre, while the discriminator is used to judge the authenticity of the generated timbre and perform feedback optimization. The generator uses a multi-layer convolutional and deconvolutional network structure to capture the complex details and harmonic characteristics of the timbre, while the discriminator improves the ability to evaluate the authenticity of the spectrogram through a multi-scale discrimination mechanism.

In order to further improve the details and authenticity of the timbre generation, this study combines data-driven methods with physical modeling technology. Specifically, the physical modeling part simulates the physical properties of the pipa, such as string tension, string plucking force, and vibration mode, thereby enhancing the authenticity of the generated audio. After the spectrogram is generated, it is converted into a playable high-fidelity audio signal through inverse transformation.

The evaluation indicators include subjective listening tests and objective audio feature analysis. The subjective evaluation invites professional pipa players and ordinary listeners to participate in the naturalness, expressiveness, and similarity of the generated audio to the real audio. The results show that 83% of the subjects believe that the generated audio is difficult to distinguish from the real audio. Objective evaluation used indicators such as signal-to-noise ratio (SNR), harmonic distortion rate, and dynamic range to verify the technical performance of the generated audio. The results showed that the SNR reached 35.2 dB and the harmonic distortion rate was less than 3%.

In addition, in the experiment of the classic pipa piece "Ten Sides of Ambush", the model successfully reproduced the timbre characteristics and playing styles of different passages, especially in the fast finger roll and complex string bending passages. It showed excellent timbre restoration ability. This generative model has high robustness in dynamic changes, harmonic details, and rhythmic consistency. This provides a good foundation for further expansion to other complex pieces in subsequent research, and brings new possibilities for the development of virtual instruments and the digital protection of traditional music.

2. Experiments and Results

2.1. Technique classification results

The experimental results show that the overall accuracy of the technique classification model on the test set reached 92.5%. Among them, the classification of fingering and harmonics performed best, with accuracy rates of 96.8% and 94.2%, respectively. This shows that the model performs well in capturing the unique features associated with these techniques. However, due to the overlap of bending and plucking in the feature space, their classification accuracy is slightly lower than other features, at 89.5% and 87.3%, respectively.

Through mixed matrix analysis, the research team identified the main feature sources of model classification and optimized them using feature enhancement techniques. For example, the sample size of a specific technique was increased, and the literature charts were enhanced to highlight their key features, thereby further improving the model's ability to distinguish easily confused categories. Table 1 summarizes the classification results of different techniques.

Table 1. Classification results of different techniques

Technique category	Classification accuracy (%)
Fingering	96.8
Overtones	94.2
Pitch bend	89.5

2.2. Tone Simulation Effects

The audio generated by the timbre simulation model has a high degree of similarity with the real audio in the spectrum graph. Subjective listening tests show that 83% of the subjects think that the generated timbre is difficult to distinguish from the real timbre. In particular, in the experiment of the classic pipa piece "Ten Sides of Ambush", the model successfully restored the timbre characteristics and playing style of different sections.

In the fast finger roll section, the generated audio can accurately simulate the clarity and continuity of the string vibration, while in the complex string bending section, the dynamic changes and harmonic characteristics of the timbre are fully restored. The results of objective evaluation also further prove the performance of the timbre simulation model, with an average signal-to-noise ratio (SNR) of 35.2 dB and a harmonic distortion rate of less than 3%, indicating that the generated audio has high fidelity.Table 2 objective indicator analysis further verifies the quality of the generated audio.

Table 2. Audio quality

Index	Value
Signal-to-noise ratio (SNR)	35.2 dB
Harmonic distortion rate	< 3%

2.3. Case Study

Taking "House of Flying Daggers" as an example, this study analyzed and simulated various techniques in its performance. The results show that the model still maintains a high degree of timbre restoration when processing fast finger rolling and complex string bending, and the rhythm consistency and dynamic changes of the audio output reach a high level.

In the simulation of fast finger rolling, the generated audio can clearly reproduce the continuity when switching fingerings, and is highly consistent with the rhythm of the real audio. In the passage of complex string bending, the generated audio shows the influence of mechanical properties on timbre, such as harmonic shift and dynamic pitch changes caused by changes in string tension. In the case of rapid switching of multiple techniques, the generated audio can smoothly connect the transient characteristics of different techniques, showing a high level of artistic expression as a whole.

Through in-depth analysis of the case, the model's excellent ability in timbre restoration, rhythm consistency and dynamic detail processing is verified, providing a solid foundation for subsequent research.

3. Discuss

3.1. Research Contributions

This study combines deep learning [4] and physical modeling technology to develop a comprehensive framework for pipa performance analysis and timbre simulation. Experimental results show that this method can not only accurately classify pipa performance techniques, but also generate high-quality timbre simulations. This provides technical support for the digital protection and dissemination of pipa art, and brings new possibilities to music education, virtual instrument development and other fields.

3.2. Limitations and directions for improvement

First, the size of the dataset is limited, which may affect the model's ability to generalize to complex repertoires and non-mainstream techniques; second, there is still room for improvement in timbre simulation in terms of extreme volume and high-frequency details. Future research can improve the robustness of the model by expanding the multimodal dataset to cover more playing styles and techniques. In addition, exploring simulation methods for multi-instrument collaborative performance and applying timbre simulation to virtual reality and gaming are also areas worthy of attention.

4. Conclusion

This study successfully developed a complete analysis and simulation framework by combining artificial intelligence technology with pipa art. The high performance of the technique classification model and the timbre simulation model verified the potential of artificial intelligence in traditional music research. In the future, further optimization and promotion of this method is expected to inject new impetus into the inheritance and innovation of pipa art, while promoting the dissemination and development of traditional music art around the world.

References

[1] McFee B. et al. (2015). "Librosa: Audio and music signal analysis in Python." Proceedings of the 14th Python in Science Conference..

[2] Li Xiangting. “Introduction to Pipa Art,” in 2015 Proc. Beijing: Central Conservatory of Music Press.

[3] Wang Cizhao. “Artificial Intelligence and Music: Theory and Application,” in 2018 Proc. Shanghai: Fudan University Press.

[4] Goodfellow I, Bengio Y, and Courville A. (2016). Deep Learning. MIT Press.