Research on Timbre Modeling and Cultural Symbol Transmission of Chinese Folk Instrumental Music Driven by Artificial Intelligence

Rui Liu

Department of Music, School of Humanities, Southwest Jiaotong University, Chengdu, Sichuan, China

Keywords: Artificial Intelligence; Folk Instrumental Music; Timbre Modeling; Cultural Symbols; Deep Learning

Abstract: Against the backdrop of the rapid rise of artificial intelligence technology, folk instrumental music, as the core carrier of Chinese traditional culture, is facing the dual challenges of digital modeling and cultural expression. This study focuses on the timbre of Chinese folk instrumental music, exploring the application mechanism of artificial intelligence in high-precision timbre modeling and cultural symbol transmission. By analyzing the limitations of current timbre modeling technology and the complexity of the timbre characteristics of folk instrumental music, deep learning, graph neural networks and cultural knowledge graphs are combined to build a modeling system that integrates timbre recognition with cultural semantic embedding. The study formulates measures such as building a high-quality timbre database, accessing cultural feedback loops, and achieving a sound-meaning mutual promotion generation system to promote the digitalization path of folk music where technology and culture are integrated and coexist. This study not only makes breakthroughs in the field of timbre modeling technology, but also provides innovative examples in the logic of cultural inheritance, opening up the application boundaries of artificial intelligence in the protection and dissemination of intangible cultural heritage.

1. Introduction

With the leapfrog progress of artificial intelligence technology, especially deep learning, generative models and audio recognition algorithms, the field of music technology has encountered unprecedented opportunities for innovation. As an important part of Chinese traditional music culture, folk instrumental music, with its unique timbre, playing techniques and cultural context, constitutes an irreplaceable sound symbol system [1]. At present, folk instrumental music is still in its infancy in the field of digital modeling and intelligent reproduction, which is particularly manifested in the lack of accuracy in timbre modeling, the potential risk of cultural semantics being out of domain, and the weakening of symbol system expression. In the context of the continuous extension of digital communication, traditional folk music is in a realistic danger of being simplified, misinterpreted or even marginalized [2]. Carrying out research on folk instrumental timbre modeling and cultural symbol transmission based on artificial intelligence can not only promote the in-depth application of intelligent music technology in the field of intangible cultural heritage protection, but also build an innovative model of synergy and mutual promotion between technology and culture, adapting to the urgent requirements of the transformation of the national cultural inheritance model in the new era. This research has a role in expanding the cross-application system of artificial intelligence in the fields of music and humanities, enhancing the scope of dissemination and international influence of folk music, and has theoretical and practical value that cannot be ignored [3].

From a deeper perspective, the application of artificial intelligence to timbre modeling should not only focus on the optimization of technical dimensions, but also introduce cultural understanding and semantic analysis mechanisms when designing models, so as to complete the deep creation of timbre connotations. The timbre of national musical instruments not only involves the physical waveform of sound, but also carries a lot of historical memories, social environment and aesthetic standards. In Jiangnan culture, the flute often symbolizes tranquility and elegance, while suona is often used in wedding and funeral ceremonies, showing distinct social contextual connections and symbolic meanings. Most audio modeling methods focus on spectral features and time domain structures, and

  DOI: 10.25236/icfmhss.2025.010

do not consider the dimensional level of these cultural symbols, which makes the generation system often show the limitation of "similar in form but empty in meaning" [4]. It is necessary to integrate the semiotic perspective into the model and create an interactive mechanism between timbre and cultural semantics. For example, cultural knowledge graphs, performance technique corpora and context labels can be used to enhance the semantic embedding ability of the model, ensure the fidelity of timbre generation in the acoustic dimension, and have expressiveness and precise directionality in the cultural field [5]. The digital protection of national instrumental music should not be limited to documentary archives and audio preservation, but should be comprehensively updated in terms of generation mechanism, interaction form and dissemination path. Artificial intelligence system can achieve this by building a cross-modal interaction model, allowing users to use text, images and even cultural keywords as input items, and the system can produce timbre fragments with corresponding cultural semantics, forming an innovative creation and dissemination model of "using culture to control timbre". This not only provides a technical guarantee for the living preservation of national music, but also extends its application boundaries in education, film and television, games and other fields, enhancing the expression and vitality of national culture in the media ecology of the new era.

2. The dilemma of modeling the timbre of ethnic instrumental music in reality and the bottleneck of cultural expression

2.1 The tension between timbral complexity and digital modeling

The sound characteristics highlighted by the timbre of Chinese folk instruments are derived from the multi-dimensional integration of culture, skills and physical properties. The source of instrumental timbre is not only the material and structure of the instrument itself, but also significantly affected by factors such as the performer's technical presentation, emotional expression, regional charm and performance environment. This extremely high complexity makes the timbre of folk instruments show dynamic nonlinearity, multi-level and context-sensitive characteristics, which creates obvious tension with the stable, quantifiable and reconfigurable information expression mode required for digital modeling . In the context of traditional artificial intelligence audio modeling, sound features are generally extracted through spectrum analysis, Mel-frequency cepstral coefficients (MFCC) and other methods [6]. Most of these methods focus on the surface physical characteristics of sound and find it difficult to grasp the cultural and emotional aspects hidden in the timbre. As a result, when reproducing the timbre of folk instruments, the model often shows a lack of layering and ethnic recognition. The playing logic and timbre dimensions of different national musical instruments are very different. For example, the different changes in erhu string vibrato, the reasonable control of flute airflow, and the guzheng glissando playing techniques all put forward higher-level requirements for the timbre generation model. How to retain the cultural rhythm and dynamic tension of the timbre during the modeling process has become a key obstacle to the application of artificial intelligence technology in the field of national musical instruments.

2.2 Shortcomings of existing modeling techniques

At present, timbre modeling technology has made certain achievements in general music generation, but it still performs poorly in terms of accurate restoration of the timbre of national instruments and cultural embedding. Most modeling systems still rely on standard sound source libraries and modeling parameters centered on the Western orchestral system, and lack structured data support for national instruments, which makes it difficult for the model to easily obtain diversified timbre inputs during the training phase. Mainstream modeling frameworks such as WaveNet , GAN-based audio generation models, and Transformer structures can process high-dimensional time series data, but they often miss the details unique to national instruments, such as delicate pitch sliding, instantaneous changes, and timing adjustments caused by emotions. The "generalization tendency" of this technical path causes the generated results to become homogenized and lack cultural specificity[7]. The existing timbre model is relatively weak in semantic modeling

and fails to achieve an effective correspondence between timbre and cultural semantic labels, resulting in the omission or misjudgment of cultural semantic information in the music generation or reproduction process. In the relevant context of traditional music, a specific timbre may symbolize sadness, joy or ritual connotations. However, if the model only treats it as a "sound spectrum", it will cause the loss of cultural functions. Improving the semantic comprehension ability of the modeling algorithm and strengthening the ability to grasp the internal logic of national instruments are necessary directions for the in-depth application of current modeling technology[8].

2.3 The Cultural "Symbolization" Crisis Faced by National Instrumental Music in the Context of Artificial Intelligence

In the music generation and dissemination system dominated by artificial intelligence, the timbre of folk instrumental music is in the stage of "symbolization" translation, that is, the timbre is extracted from its original context and transformed into a callable and copyable digital resource. Although this translation is conducive to improving the efficiency of dissemination and the convenience of use, it actually triggers the hidden danger of the hollowing out of cultural semantics in terms of essential connotation. The sound of folk instrumental music is not just a physical phenomenon, but also a cultural symbol that carries ethnic memory, regional spirit and historical experience. The generation of this meaning depends on the specific performance venue, ritual structure and cultural coding of the audience[9]. When the timbre is processed into AI training samples or materials, its cultural context is often ignored or even simplified into the so-called "sound style" or "national color", which causes the timbre to no longer belong to the category of cultural behavior of embodied experience, but instead become a sound symbol resource at the visual and parameter level. This kind of cultural "decontextualization" not only weakens the cultural orientation of timbre, but also causes misinterpretation and alienation of AI-generated music in the field of semantic communication. When modeling algorithms, they mostly tend to "average timbre" or "optimal solution" based on big data optimization logic, which aggravates the marginalization of ethnic instrumental music and non-mainstream performance techniques, and accelerates the internal contraction process of the traditional music ecology. In the modeling practice where artificial intelligence plays a leading role, how to introduce cultural reproduction mechanisms, build context response models, and pay attention to the cultural coding of both performance and reception are the core paths to avoid the crisis of "symbolization" of national instrumental music.

3. The dual mechanism of artificial intelligence on the timbre of national instrumental music and cultural symbol system

3.1 Application path of deep learning models for timbre modeling

The introduction of deep learning technology has endowed timbre modeling with powerful data processing and feature learning capabilities, and has become the main approach to audio modeling. In the work of modeling the timbre of folk instruments, the models commonly used include convolutional neural networks (CNN), recurrent neural networks (RNN), self-attention mechanisms (Self-Attention) and Transformer architectures. These models can deeply abstract features such as the spectrogram, Mel-frequency cepstral coefficients (MFCC), and constant Q transform (CQT) of the original audio signal to achieve modeling of dynamic changes in timbre [10]. CNN is very good at processing local time-frequency features, which is conducive to capturing local performance techniques commonly seen in folk instruments; while RNN and Transformer show certain advantages in modeling the time series structure of timbre, and can learn the long-term dependency associations of performance techniques. In recent years, the emerging variational autoencoders (VAE) and generative adversarial networks (GANs) have been widely used in tasks related to timbre generation to construct a latent space to achieve continuous presentation and style conversion of timbre. Even though the above models perform outstandingly in the modeling of complex timbre, they still encounter problems such as scarce samples, lack of style diversity, and lack of cultural semantics in the application scenarios of folk instrumental music. To build a deep learning model, it

is necessary to combine a culturally targeted data preprocessing mechanism and a structural optimization strategy in order to achieve the preservation of timbre semantics and style-guided generation .

Table 1 Comparison of the performance of current mainstream timbre modeling methods in the task of ethnic instrumental audio recognition

Model Name	Accuracy	F1-Score	Processing time (seconds/segment)	Tone detail retention score (out of 10)
CNN	84.20%	0.812	0.23	6.5
RNN (LSTM)	85.90%	0.827	0.45	7.1
Transformer	88.50%	0.854	0.39	8.2
VAE + CNN	87.30%	0.841	0.31	7.8
GAN (tone generation)	81.40%	0.793	0.51	6.3

As can be seen from Table 1, the Transformer model has outstanding advantages in overall recognition accuracy and F1-Score, especially in preserving timbre details. Even though the RNN model has certain advantages in modeling dynamic changes, its processing time is longer and its application scope is limited. GAN has shown certain timbre generation capabilities, but due to the instability of training, its recognition performance is relatively poor. The model formed by the fusion of VAE and CNN has achieved a good balance between recognition accuracy and timbre details, and is suitable for applications that compromise between accuracy and generation efficiency. This table supports the selection of models, not only considering accuracy, but also considering the ability to restore timbre details that cultural characteristics rely on.

3.2 Extraction and semantic encoding of timbre features of ethnic instrumental music

The general audio task is very different from the extraction of timbre features of folk instruments. It not only pays attention to the physical properties of the sound, but more importantly, it is about how to reveal the cultural semantics and technical characteristics behind it. In the process of timbre modeling, it is necessary to extract multi-dimensional features including pitch envelope, resonance peak structure, attack duration, time domain instability, noise ratio, etc. from the audio signal, so as to accurately reflect the powerful explosiveness of suona, the exquisite and delicate glissando of erhu, and the unique airflow resonance of flute. These low-level features need to be coupled with high-level semantic structures to achieve effective semantic encoding [11]. Tone modeling has begun to adopt multimodal learning methods to embed audio signals with text labels, image context or music score data into a unified vector space, relying on alignment mechanisms to enhance the semantic association effect. In the work of modeling ethnic musical instruments, this can be achieved by introducing regional labels, playing technique markers, cultural context classification, etc., so that the model can simultaneously explore the mapping relationship between timbre and cultural semantics during the training process. Such semantic encoding is beneficial to improving the classification and recognition accuracy of the model, and further provides structured guarantees for subsequent timbre generation, retrieval and cultural communication activities. The deep combination of feature extraction and semantic modeling has become the key supporting architecture for the intelligent modeling of ethnic musical instrument timbre.

3.3 The Coding Logic of Timbre as a Cultural Symbol

As a cultural symbol, timbre is not an isolated single sound form, but a complex coding system that encompasses cultural knowledge, historical experience and social context. In the context of folk music, specific timbre often carries conventional cultural connotations, such as the emotional symbols embodied by suona during wedding and funeral ceremonies, the ethereal atmosphere conveyed by the timbre of the flute, and the classical aesthetic associations evoked by the glissando of the guzheng. The cultural symbolic attributes of timbre originate from the construction of community cognition and perception structure in long-term social interactions . Its generation and transmission are carried out under the joint influence of historical context, ritual structure,

performance style and audience experience. Timbre is by no means just a "set of sound features" at the technical level, but also a messenger that carries and conveys cultural significance [12]. During the artificial intelligence modeling work, if the cultural symbolic connotation of timbre is omitted and only the numerical modeling of physical signals is carried out, it is easy to cause the semantics to lose its original context and the culture to become flat. In order to achieve the goal of timbre encoding that embeds culture, it is necessary to introduce an intermediary mechanism between timbre and cultural semantics, such as building a knowledge graph of timbre-technique-scene-emotion, or constructing a multi-relationship symbol propagation model based on a graph neural network. By relying on the symbolic logic model embedded in the timbre of national instruments, timbre modeling can be promoted from the perception of low-level features to the construction of high-level culture, thereby realizing the semantic recognition and cultural inheritance efficiency of artificial intelligence in national music.

4. A method for achieving timbre modeling and cultural symbol co-construction based on artificial intelligence

4.1 Building a high-quality folk instrumental music database

As the infrastructure for intelligent modeling of national musical instruments, a high-quality timbre database is the foundation. Its construction level directly affects the training effect, generation capacity and depth of cultural expression of the model. In the field of national musical instruments, the construction of the database should meet the comprehensive standards of multi-level, high-fidelity and cultural diversity. The sampling data should cover the representative national musical instruments. The recording should be carried out in an environment without echoes, and high-resolution audio format should be used to capture the details of complex timbre. Systematic timbre data collection of various playing techniques should be carried out, involving playing techniques such as glissando, vibrato, tremolo, beating and tonguing. It is also necessary to integrate the different styles of different performers and the level of playing intensity, so as to expand the performance dimension of the database. The corresponding information of the instrument timbre and social and cultural background should be embedded, such as regional affiliation, context of use, historical function and emotional connotation, so as to achieve the semantic labeling of data. In terms of the organization and arrangement of data, it is advocated to adopt a multimodal database model, combining audio, music scores, text annotations, image data and video performances, to achieve collaborative training of timbre modeling and cultural semantics, and adopt overall improvements at the data level. It can provide deep learning models with an accurate, rich, and culturally sensitive training foundation, and build a knowledge context and application foundation for intelligent timbre modeling (See Table 2).

Table 2 Sample coverage and cultural semantic label distribution statistics (based on the prototype of the national instrumental music database)

Musical instrument category	Number of samples (items)	Number of techniques involved	Number of cultural semantic tags	Coverage area
Guzheng	3,240	12	9	Jiangnan, Lingnan, North
Erhu	2,710	10	8	North China, Jiangnan
Flute	2,980	11	10	Sichuan, Chongqing, Yunnan, Guizhou, and Southern Fujian
Suona	2,320	9	7	Central Plains, Northeast China
lute	2,680	13	10	Jiangsu, Fujian, Sichuan

The database covers five categories of core national musical instruments and 930 high-specification timbre samples, comprehensively covering playing techniques and numerous regional styles. Guzheng and pipa are used as examples. Both have large sample sizes and a wide range of playing techniques, demonstrating their ability to present complex playing features in modeling.

Cultural semantic labels cover a very wide range, including both ritual functions and regional cultural characteristics, demonstrating that the database has good basic cross-modal modeling capabilities. This type of structured data resource provides a lot of semantic support for the subsequent construction of graph neural networks and knowledge graphs, and is helpful in promoting the transformation of timbre from "physical layer sampling" to "cultural symbol embedding."

4.2 Introducing attention mechanism and graph neural network to enhance the semantic depth of modeling

At present, one of the main challenges facing timbre modeling technology is that it is difficult for the model to grasp the complex mapping relationship between timbre and its cultural semantics. The introduction of attention mechanism and graph neural network provides a technical breakthrough for enhancing the depth of semantic modeling. The attention mechanism relies on the operation of dynamically allocating weights to allow the model to focus on the key segments with the most significant cultural expression in time-series audio, which is particularly suitable for the sound characteristics of folk instruments such as glissando, staccato, and free rhythm, which have significant representational significance, and enhances the model's perception of local subtleties of timbre. Based on this foundation, graph neural networks (GNNs) can process multi-dimensional correlation structures and are suitable for constructing graph structures between timbre and playing techniques, cultural scenes, and instrument types. They can embed timbre into a knowledge space with semantic connections for dissemination and evolution. By combining the attention mechanism with graph neural networks, it can achieve feature learning from the bottom layer and semantic supervision from the top layer. In the modeling process, physical feature extraction and cultural label prediction can be carried out simultaneously to overcome the existing problems of semantic ambiguity and symbol delocalization in timbre modeling. This type of integrated model not only improves the cultural relevance and recognition of timbre generation, but also provides structured support for music intelligence systems in cross-context applications.

4.3 Design of the timbre generation system based on cultural knowledge graph

The cultural knowledge graph builds a key intermediary bridge for the semantic expression and generation control of the timbre of national instrumental music. The knowledge graph can systematically connect the timbre with its related cultural concepts, performance techniques, historical scenes, and emotional semantics by constructing entity-relationship-attribute triples, so as to achieve contextual control and semantic guidance of timbre generation. Introducing the cultural knowledge graph into the timbre generation system, firstly, the cultural nodes of the graph can be mapped into the potential space of the timbre model by means of semantic embedding technology, so as to realize the construction of an integrated generation path from "cultural intention" to "timbre output"; the graph convolutional neural network (GCN) or graph attention network (GAT) can be used to encode the graph, so as to enable the system to understand the correlation strength of each cultural element, so as to automatically adjust the matching of timbre style and cultural semantics during generation. The system also supports interactive generation mode, allowing users to input cultural labels such as "Tibetan wedding" and "Jiangnan Water Village". The system will generate corresponding timbre fragments accordingly, achieving semantic mapping and transformation of timbre and cultural symbols, and introducing cultural knowledge graphs. This not only enhances the explanatory power and controllability of the timbre modeling system, but also promotes the deep transformation of artificial intelligence music generation technology from data-driven to knowledge-driven.

5. Conclusion

This paper conducts a systematic study on the artificial intelligence modeling and cultural symbol transmission of Chinese folk instrumental timbre, develops a modeling path with technical methods and cultural meaning as the core, adopts deep learning and graph neural network, realizes high-precision modeling of the timbre characteristics of various folk instruments, breaks the dilemma of

existing technology in dynamic timbre expression and semantic understanding; by building a cultural knowledge map, introducing performance context and regional cultural elements, it achieves cultural integration in the timbre generation link, and greatly enhances the symbolic level of the model. Digital modeling of folk instruments is not just about audio reconstruction, but also focuses on the generation and dissemination of culture. In this process, artificial intelligence technology should remove the role of neutral tools and become an active actor in the cultural co-creation mechanism. Future research should further expand the scope of interdisciplinary collaboration, create a more open, diverse and culturally sensitive training system, and enhance the public's voice in technical modeling and cultural selection. This study provides a theoretical basis and practical framework for the digital protection, generation and dissemination of folk music, and has significant academic value and social significance.

References

[1] Chen D, Sun N, Lee JH, et al. Digital Technology in Cultural Heritage: Construction and Evaluation Methods of AI-Based Ethnic Music Dataset[J]. Applied Sciences, 2024, 14(23): 10811.

[2] Morales GS, Perez MLAC, Tabuena A C. Artificial Intelligence and the Integration of the Industrial Revolution 6.0 in Ethnomusicology: Demands, Interventions and Implications[J]. Musicologist, 2024, 8(1): 75-107.

[3] Mehta A, Chauhan S, Choudhury M. Missing Melodies: AI Music Generation and its" Nearly" Complete Omission of the Global South[J]. arXiv preprint arXiv:2412.04100, 2024.

[4] Hui F. Transforming educational approaches by integrating ethnic music and ecosystems through RNN-based extraction[J]. Soft Computing, 2023, 27(24): 19143-19158.

[5] Pi J. Application of Artificial Intelligence Technology in Music Creation [J]. Musical Instrument, 2025, (09): 68-71.

[6] Aung K K , Nakabayashi Y, Shioya R, et al. A Comparative Study of Generative LSTM Models for Multi-Instrumental Music Composition[J]. International Journal of Computer Science & Network Security, 2025, 25(4): 11-26.

[7] Chen M, Xiang Y, Xiong C. Synthesis and Restoration of Traditional Ethnic Musical Instrument Timbres Based on Time-Frequency Analysis[J]. Traitement du Signal, 2024, 41(2).

[8] Wen YW, Ting C K. Recent advances of computational intelligence techniques for composing music[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022, 7(2): 578-597.

[9] Wang L, Zhao Z, Liu H, et al. A review of intelligent music generation systems[J]. Neural Computing and Applications, 2024, 36(12): 6381-6401.

[10] Berkowitz A E. Artificial intelligence and musicking: A philosophical inquiry[J]. Music Perception: An Interdisciplinary Journal, 2024, 41(5): 393-412.

[11] Lu N. Research on Intelligent Assistant Teaching System for National Musical Instruments Based on PSS[M]//Design Studies and Intelligence Engineering. IOS Press, 2024: 1083-1095.

[12] Clark E H. A Postcolonial MIR?[J]. Resonance: The Journal of Sound and Culture, 2022, 3(4): 412-432.