ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Multi-dimensional Approaches to the Protection and Inheritance of Traditional Music Driven by Artificial Intelligence

Jiangzhao Ye[1,*]

[1] Jeonbuk National University, Jeonju, 54896, Korea

SUMMARY: *As a unique art form, traditional music protection is facing increasingly severe challenges as the times change, and traditional music inheritance is becoming more and more difficult. Based on analyzing the contemporary value of traditional music, the article describes the specific application of artificial intelligence technology in the protection and inheritance of traditional music. In order to realize the innovative development of traditional music, this article designs the CT Music database. Then the random mask is combined with Transformer to construct RM-Transformer model for generating traditional music chords, and GCT model is designed for traditional music generation. The results show that compared with the mainstream model, the number of parameters and the average time of generating music of the GCT model are reduced by 0.98MB and 3.22s, respectively, and the scores of the generated traditional music are all higher than 34.5 on the subjective evaluation of listening. The full application of artificial intelligence technology can significantly improve the generation quality of traditional music and provide a development path for the protection and inheritance of traditional music.*

KEYWORDS: *traditional music; CT Music database; Transformer; GCT model; music generation*

1 Introduction

As an important part of the excellent traditional Chinese culture, traditional Chinese music is not only a form of musical art expression and carrier, but also carries the history, spirit, emotion, social production and other aspects of the nation and the country, and the protection and inheritance of traditional Chinese music has extremely important value of the times, historical value and cultural value [1-4]. With the rapid development of society and the strengthening of international cultural exchanges, the introduction of western music culture has had a certain impact on Chinese traditional music culture [5, 6]. Especially with the change of social pattern and the change of music culture ecological environment, Chinese traditional music is in a disadvantageous position in the competition with foreign music culture, gradually marginalized, and even facing the dilemma of gradual disappearance [7-9]. Therefore, we must recognize the importance and urgency of the protection and inheritance work of traditional music culture.

Artificial intelligence, as a rapidly developing frontier technology in recent years, has shown great potential and application prospects in various fields [10]. In the protection and inheritance of traditional music, AI technology can provide us with new solutions and bring great innovation and help to the protection and inheritance of traditional music [11, 12]. Driven by artificial intelligence, traditional music protection and inheritance is mainly realized through

*18758121764@163.com
https://doi.org/10.65102/is2026612
Published: 30 April, 2026

intelligent creation, immersive experience and other multi-dimensional methods. By analyzing a large number of traditional music works and artists' musical styles, AI can generate compositional fragments with different stylistic characteristics [13, 14]. At the same time, AI can simulate human creativity and imagination to create traditional music works with stunning effects, which can not only bring people new feelings in terms of aesthetics, but also help traditional music creators to further deepen their understanding and exploration of music creation [15-18]. And the combination of traditional music and virtual reality opens a new chapter of immersive listening experience [19]. The application of virtual reality technology enables traditional music to break through the limitations of physical space and create an immersive listening environment, in which every note seems to ring in the ear, and every piece of music can cause deep emotional resonance in the listener, which is more conducive to the protection and inheritance of traditional music [20-23].

In order to expand the innovative path of traditional music protection and inheritance, the article proposes a traditional music generation GCT model based on Transformer network. The model fully considers the correlation between melody and chord in traditional music, and realizes the generation of traditional music melody with the help of RM-Transformer model.The GCT model further reduces the number of parameters, significantly improves the computational efficiency, and the generation quality of traditional music increases significantly.

2 Traditional music database construction

Artificial Intelligence (AI), as a “new quality productivity” driving social change, is deeply reshaping the underlying logic and ecological structure of traditional music preservation and inheritance. In the field of traditional music preservation and inheritance, this technological wave has transcended the instrumental auxiliary level, and has had a subversive impact on the core links of music creation, performance, dissemination and acceptance. In this context, sticking to the traditional mode is not only difficult to adapt to the acceleration of technological iteration, but also may lead to the disconnection between traditional music and social needs, lose its attraction to the new generation of learners, and eventually fall into an existential crisis of systemic failure.

2.1 Artificial Intelligence and Traditional Music

2.1.1 Contemporary value of traditional music

(1) Enhancing a sense of identity and pride. In today's increasingly globalized economy, traditional music, as a carrier of national culture and historical memory, can help people understand and pass on their own cultural traditions and enhance their sense of national identity and pride. It helps to maintain the diversity and uniqueness of national culture and promote cultural exchange and integration.

(2) Enhance artistic aesthetic ability. As a form of art, traditional music has unique aesthetic value. With its unique melody, harmony and rhythm, it triggers people's deep emotional resonance.

(3) Improve the international influence of Chinese culture. As an important part of Chinese culture, traditional music has a unique charm and attraction, and has wide influence in the international arena. Through the dissemination of Chinese traditional music, understanding and communication between different cultures can be strengthened, and while promoting cultural exchanges and mutual understanding among countries around the world, it also helps to enrich the world's music culture and increase the international influence of Chinese culture.

2.1.2 Practical applications of AI technology

The integration and development of artificial intelligence technology and traditional music has greatly innovated the form of music dissemination and popularization, so that the path of music dissemination has been transformed from the primitive, unidirectional oral and ear-to-ear transmission to a multidimensional, diversified dissemination that is not subject to the limitations of geography and time and space. The application of artificial intelligence in the field of traditional music has brought a new artistic experience to the public.

(1) Melody Recognition. Compared to the retrieval of a given audio file, AI intelligent recognition does not need to carry out tedious steps such as format conversion and file input, and can directly recognize the music melody based on the playback. When the intelligent device receives a certain melody signal, it will apply computer algorithms to screen and recognize it, and determine whether the melody segment exists in the device's existing music data resource library. Then the pitch, rhythm and other signals in the resource library will be extracted, matched with the existing melody, and finally arrive at the answer.

(2) Intelligent Composition: The AI composition function not only simplifies the complexity of composition, but also enhances the universality of traditional music creation.The AI composition method forms a neural network inside the computer with the help of a huge amount of songs, lyrics, tunes and other musical elements. It then automatically arranges harmonies according to the user's needs, and ultimately outputs a privately customized musical composition.

2.2 Traditional Music Database Construction

2.2.1 Data collection steps

In order to fully capture and translate traditional music, the acquisition needs to systematically cover the pitch, timbre and playing techniques of all instruments of traditional music. Figure 1 shows the steps of traditional music data acquisition as follows:

(1) Process the collected traditional music mix recording files by AI instrument recognition technology, identify the corresponding instrument timbre of each voice part, and separate them into independent audio tracks.

(2) Convert the separated audio tracks into Mel spectra, which is a visual representation of the audio spectrum that is more suitable for processing and analysis by machine learning algorithms.

(3) Generate raw MIDI sequence files based on the Mel Spectrum, this step can be carried out automatically by machine learning algorithms to convert the note information in the Mel Spectrum into note sequences in MIDI format. As for the recording files that have been split into tracks, they can be directly converted to MIDI format to simplify the process of transcription.

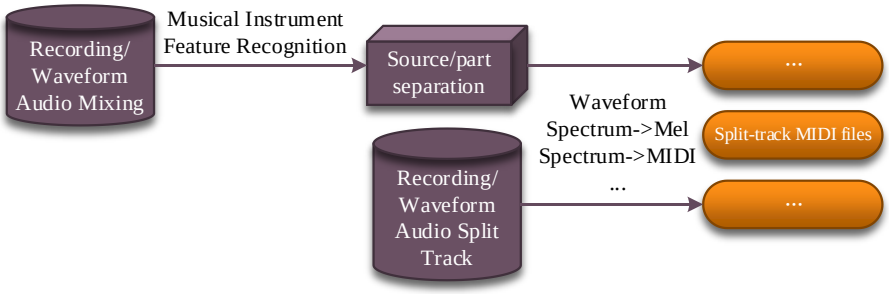

Figure 1: Traditional music data collection steps

2.2.2 CT Music database

Based on the traditional music data collection steps, this paper selects a number of traditional music to construct the CT Music database, which contains a total of 40 songs (more music data are being labeled), all of which are Chinese traditional music. For each song in the CT Music database, the labeling information is divided into four categories, i.e., meta-information, expert subjective evaluation, vocal track labeling information, and accompaniment track labeling information.

(1) Song meta-information labeling. CT Music database meta-information includes song name, singer's name, singer's gender, song length, song tempo, song language, song beat, and song key number.

(2) Vocal track information labeling. Database vocal track annotation includes song singing style, song timbre, song synchronization with lyrics, beat time start point (in s), bar line start and end time (in s), language, timestamp of the 1st tone of each bar (in s), and frequency of the 1st tone of each bar (in Hz). The timbre of the song was labeled into six categories, i.e., thick, raspy, powerful, sweet, ethereal, and high, which can be used for the timbre recognition task.

(3) Subjective high-level evaluation labeling. The subjective advanced evaluation annotations include whether the emotion is full (Y/N), whether the range is appropriate (Y/N), whether the real and falsetto voices are seamlessly transformed (Y/N), whether the breath is abundant (Y/N), whether the timbre is recognizable (Y/N), and whether the diction is clear and accurate (Y/N).

(4) Accompaniment track information labeling. Accompaniment track information labeling includes beat time points, bar lines, chords, timing of the first note of each bar (in s), and instruments used.

3 Music generation model integrating traditional music

Traditional music is an artistic treasure in the treasure house of human culture, which not only carries the memory of history, but also embodies the characteristics and emotions of the Chinese nation. Chinese traditional music contains many different types of expression, and these distinctive forms of music together constitute a rich and colorful traditional music culture, which is an important carrier for the cultural heritage and national identity of the Chinese nation. However, due to geographical differences and imbalance of resources, traditional music protection and inheritance have always faced the real problems of lack of resources and relatively single mode, which are difficult to meet the needs of traditional music inheritance.The emergence of AI has brought new opportunities and challenges for traditional music protection and inheritance.

3.1 Attention Mechanisms and the Transformer Model

3.1.1 Multi-attention mechanisms

The multi-head self-attention mechanism is an attention mechanism for processing sequential data. Compared to the traditional self-attention mechanism, the multi-head self-attention mechanism has more powerful modeling capabilities because it allows the model to simultaneously notice different relevant information in different representation subspaces. In the traditional self-attention mechanism, the model weights and aggregates the vectors at each position in the input sequence to obtain a vector representing the contextual information at that position. The multi-head self-attention mechanism, on the other hand, introduces multiple

attention heads, each of which learns a different attentional weight, thus enabling the model to simultaneously attend to different dependencies in the input sequence.

Specifically, the multi-head self-attention mechanism first maps the input sequence into multiple different subspaces, then computes the attention weights and weighted summation in each subspace separately, and finally stitches together the results from the different subspaces to obtain the overall attention representation. In the multi-head attention mechanism, x_i and x_j represent two different word/phrase embeddings in the input sequence, respectively. Then the two sets of query vectors, key vectors and value vectors are denoted as:

$$\begin{cases} q_i = W^q x_i \\ k_i = W^k x_i \\ v_i = W^v x_i \end{cases} \tag{1}$$

$$\begin{cases} q_j = W^q x_j \\ k_j = W^k x_j \\ v_j = W^v x_j \end{cases} \tag{2}$$

Among them, the query vector q_i is used to compute the correlation between the current word/phrase element x_i and other words/phrase elements, the key vector k_i is used to compute the similarity between x_i and words/phrase elements at other positions, and the value vector v_i is used to construct the contextual representation. The W^q, W^k, W^v are the weight matrices, which are updated during model training.

The multi-head self-attention mechanism assigns multiple query vectors, key vectors and value vectors to each word/word element again. Taking two attention heads as an example, the computation process of multi-head query vectors, key vectors and value vectors for x_i is as follows:

$$q_{i,1} = W^{i,1} q_i \tag{3}$$

$$q_{i,2} = W^{i,2} q_i \tag{4}$$

$$k_{i,1} = W^{k,1} k_i \tag{5}$$

$$k_{i,2} = W^{k,2} k_i \tag{6}$$

$$v_{i,1} = W^{v,1} v_i \tag{7}$$

$$v_{i,2} = W^{v,2} v_i \tag{8}$$

where $W^{i,1}, W^{i,2}, W^{k,1}, W^{k,2}, W^{v,1}, W^{v,2}$ is the weight matrix that can be updated during model training. Next, corresponding to the two attention heads, the attention weight $\hat{\alpha}_{ij}$ for each word/word element is obtained by SoftMax normalization of the attention scores, i.e:

$$s_{ij} = F\left(q_i, k_j\right) \tag{9}$$

$$\hat{\alpha}_{ij} = \frac{\exp\left(s_{ij}\right)}{\sum_j \exp\left(s_{ij}\right)} \tag{10}$$

Finally, the model splices $b_{i,1}$ and $b_{i,2}$ to obtain the contextual features b_i generated for x_i under the multi-head self-attention mechanism. Namely:

$$b_i = \sum_j \hat{\alpha}_{ij} v_j \tag{11}$$

The multi-head self-attention mechanism introduces multiple attention heads, each of which can focus on different semantic features. This helps the model to capture information from different perspectives and at different granularities at the same time, thus providing a richer representation of information. In the sequence coding task, the multi-head self-attention mechanism is able to capture features at different levels in the sequence, which helps to improve the model's representational capability. In addition, the multi-head self-attention mechanism is able to pay global attention to different locations in the sequence, thus capturing a global representation of the features. This helps to improve the generalization performance of the model, allowing the model to better handle variable-length sequences or sentences of different lengths.

3.1.2 Transformer model

Transformer is a novel neural network framework, Transformer network consists of encoder and decoder. In the encoder and decoder, Transformer captures information about interactions between different features through the mechanism of multi-head attention. The main function of the encoder is to extract information from the input sequence and encode the entities in the sequence into a representation vector of a specific length, and the decoder converts this representation vector into a sequence of entities for the target task. Both the encoder and decoder consist of several modules containing a self-attention mechanism network and a feed-forward neural network (FFN). Specifically, in the self-attention mechanism, the word vectors at each position are computed with the word vectors at their own and other positions to obtain a weighted sum, with the weights determined by the attention scores. The formula for self-attention is as follows:

$$Attention(Q, K, V) = \mathrm{SoftMax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{12}$$

where Q, K and V denote the query vector, key vector and value vector, respectively, and d_k denotes the dimension of the word vector.

In the Transformer model, Q, K and V are computed from the output of the previous layer.

In feed-forward neural network, the word vector of each position first passes through a fully connected layer, then a live number (usually ReLU function), and finally passes through another fully connected layer to get the new word vector representation. The formula for feedforward neural network is as follows:

$$FFN(x) = \max\left(0, xW_1 + b_1\right)W_2 + b_2 \tag{13}$$

where x denotes the input word vector, and W_1, b_1, W_2 and b_2 denote the weights and biases of the two fully-connected layers, respectively.

In each self-attention mechanism network layer, Transformer also incorporates residual connectivity and layer normalization, which can improve the training accuracy and robustness of the model.

3.2 GCT-based music generation modeling

3.2.1 The RM-Transformer Model

In this section, an RM-Transformer encoder is designed, firstly, in the Transformer encoder method, the position encoding and memory mechanism are introduced, and axial attention is added, the combination of these methods can be effective for deep feature extraction.

Position encoding provides the model with information about each position in the sequence. Even if part of the region is masked, the position encoding is able to retain its positional information, thus enabling the model to distinguish the hidden part. This positional information is critical for the network to understand the location of the original sample. In the encoder, the position encoding is a learnable weight, of the shape (max_seq_length,key_dim), that provides positional information by summing with the input sequence. This design allows the model to capture the relative and absolute positional relationships of elements in the sequence. To wit:

$$PE(pos, 2i) = \sin\left(pos / 10000^{2i/key\_dim}\right) \tag{14}$$

$$PE(pos, 2i+1) = \cos\left(pos / 10000^{2i/key\_dim}\right) \tag{15}$$

where pos is the position and i is the dimension. In the model, the position encoding is learnable and the position encoding PE is added to the input of the Transformer encoder ($\tilde{Y}$ of the DCNN output) to provide position information. Let the input of the Transformer encoder be x. The input X_{pos} after adding the position encoding is:

$$X_{pos} = \tilde{Y} + PE \tag{16}$$

Memory mechanisms allow the model to save and utilize information from previous locations to facilitate contextual understanding, and this mechanism helps the network to remember long-term dependencies in sequences to better understand and predict patterns in sequences. When processing sequence data, the memory mechanism allows the model to access previously processed information. Suppose that at time step t, the model receives input X_{pos} and has a memory set M containing information from all previous time steps. For a simplified description, assume that the memory mechanism is internally updated as:

$$M_{new} = Update\left(M, X_{POS}\right) \tag{17}$$

Then, the axial attention mechanism is used simultaneously. The effect of axial attention is indirectly realized by combining the outputs of row polytope and column polytope along one axis (in this case, the feature or dimension axis). The role of axial attention in Transformer is to

capture the relationship between different dimensions and between different positions in the sequence data, so as to better understand the shallow and deep information of the sequence data.

Let H_{row} and H_{col} be the outputs of row attention and column attention, respectively, where $MHSA(\cdot)$ denotes a multi-head self-attention module that can process the inputs and their associated memories to provide a weighted input representation. The whole process can be described as follows:

$$H_{row} = MHSA\left(X_{pos}, M\right) + X_{pos} \tag{18}$$

$$H_{col} = MHSA\left(X_{row}, M\right) + X_{row} \tag{19}$$

The multi-head attention mechanism allows the model to learn information from different representation subspaces at the same time, and for each head, its computational expression in computing attention is:

$$Attention(Q,K,V) = \text{softmax}\left(QK^T / \sqrt{d_k}\right)V \tag{20}$$

where (Q,K,V) are the query, key, and value matrices, respectively, and d_k is the dimensionality of the key, which is used to scale the result of the dot product, avoiding excessively large values that would cause the SoftMax function to be in the saturated region and affect the gradient propagation.

Finally, a normalization layer as well as a feed-forward neural network module are used, and together these components provide the model with a powerful representation learning capability. Then:

$$O = FFN\left(H_{col}\right) + H_{col} \tag{21}$$

where H_{col} is further processed to produce the final output through a feed forward neural network (FFN)

The feed-forward network (FFN) consists of two linear layers and a nonlinear activation function $ReLU$, i.e:

$$FFN(x) = \max\left(0, xW_1 + b_1\right)W_2 + b_2 \tag{22}$$

The RM-Transformer encoder is able to process sequence data efficiently by combining multi-head attention, positional encoding, memory mechanisms, and feed-forward neural networks to capture long-distance dependencies and maintain global relationships of elements in a sequence. By learning positional encoding and utilizing memory mechanisms, it enhances its ability to understand the temporal dimension of sequences, providing powerful support for complex sequence processing tasks.

3.2.2 GCT network structure

For the traditional music generation task, the problems that need to be solved mainly include music repetition, music chord generation, etc. Chords are the finishing touch of a piece of music, and there is disharmony between the chords and the main melody when chord generation occurs, and improving the degree of harmony between the two is the most important thing.

In this regard, this paper combines the existing related research to construct a divisional combination Transformer model (GCT), whose specific framework is shown in Figure 2. The melody generator is built by the RM-Transformer model in Section 3.2.1, i.e., Random Mask Transformer, the training data in the melody generator does not involve too much chord information, and the model can better learn the musical features about the melody, such as pitch and tempo.

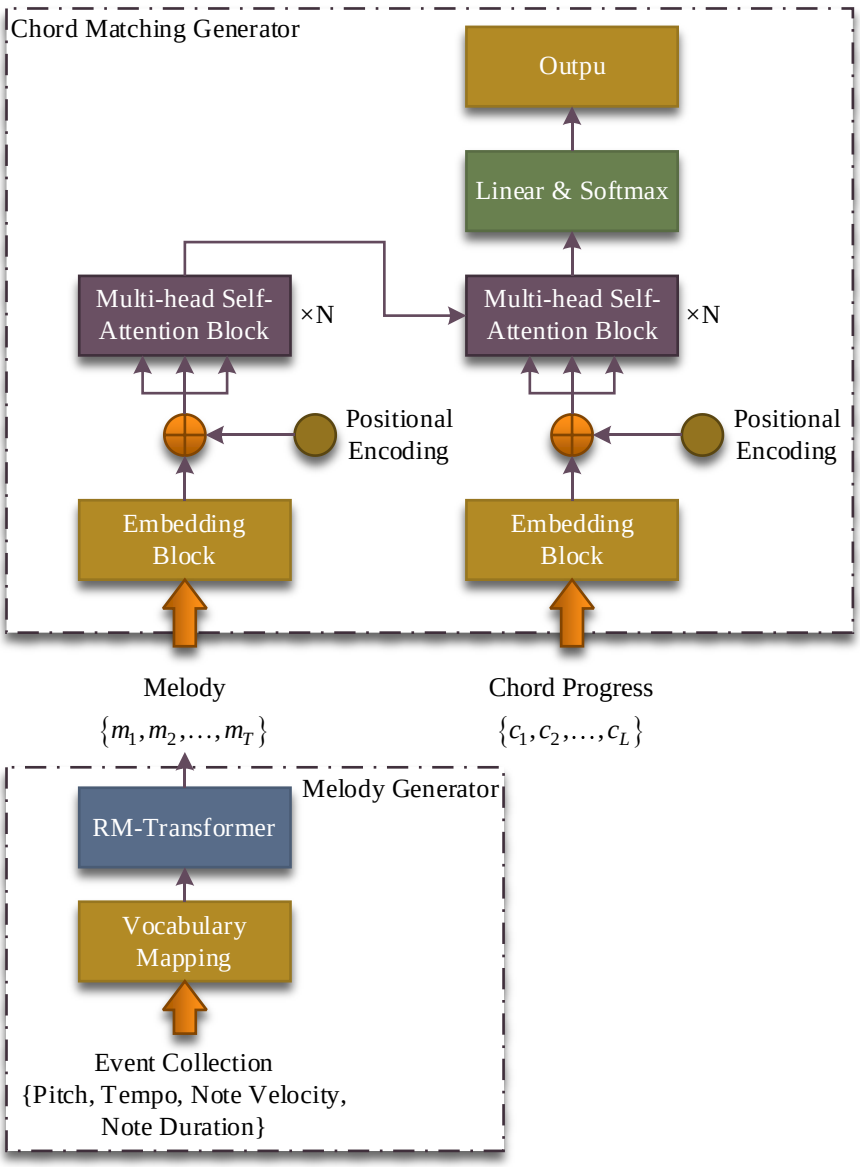

Figure 2: GCT model framework

The melody generation formula is as follows:

$$m_{1:T} = RM - Transformer\left(Melodyy_{events}\right) \tag{23}$$

where $Melody_{events}$ is the melody related events, containing note on, duration, note velocity, pitch information.

After the melody is generated into the training chord matching module, we get the melodic

sequence $\{m_1, m_2, \ldots, m_T\}$ as input to the encoder, T is the length of the melody, and the chord sequence $\{c_1, c_2, \ldots, c_L\}$ is used as the decoder input. The computation process of the chord module is as follows:

$$M_T = Embedding\left(m_{1:T}\right) \tag{24}$$

$$C_L = Embedding\left(c_{1:L}\right) \tag{25}$$

$$O_L = AttBlock\left(c_L, Self - AttBlocks\left(M_T\right)\right) \tag{26}$$

$$p\left(\tilde{y}_{1:L}\right) = Softmax\left(Linear\left(O_{1:L}\right)\right) \tag{27}$$

M_T is the feature vector after the embedding layer, the embedding layer is the normal Transformer's embedding module with position information added to this layer.Self-AttBlocks denotes the encoder module of a Transformer with S-header N layers, and the output of this encoder goes to the AttBlocks decoder, which is a decoder module of the same layer which calculates the attention of the chord and melody. The $p\left(\tilde{y}_{1:L}\right)$ is the output of the obtained event, and the final probability is estimated by a final linear layer.

In the GCT model, the chord matching generator is trained by using the database melody as the encoder input and the chords as the decoder input, after the training is completed, it can be used for melody matching and chord generation, after the melody is generated, it enters the chord matching generator to get the matching chords, and finally the complete traditional music sample with chords is obtained by the splicing of the music melody and chords.The emergence of the GCT model provides an opportunity for the application of AI to the traditional music preservation and inheritance. The emergence of GCT model provides a new research path for the application of artificial intelligence in traditional music protection and inheritance, and promotes the innovative development of traditional music inheritance.

4 Traditional music generation model validation and application

Rapid advances in artificial intelligence (AI) technology offer new possibilities for cultural preservation and innovation.AI technology, through deep learning, natural language processing, and image generation, is able to simulate and create highly artistic content. In the field of traditional music creation and dissemination, AI technology has shown great potential, capable of generating melodies, lyrics, and even simulating instrumental performance. Through AI technology, traditional music can not only be effectively preserved and inherited, but also more widely disseminated and innovated globally.

4.1 Validation of the validity of the GCT model

4.1.1 Analysis of Model Composition Results

The analysis of the results of traditional music generation was first carried out using multiple objective metrics and model comparisons, and the models compared with the models in this chapter are Polyphony RNN based on recurrent neural network and Pop music transformer

polyphonic piano auto-composition model based on Transformer network. In terms of representation, Polyphony RNN uses midi-like stream representation while Pop music transformer uses REMI representation. The conventional music generation models were trained separately using the CT Music dataset.

The first objective comparison experiment focuses on the comparison of the models in terms of their number of participants, and their composition time, and its final results are shown in Table 1. In this case, the average time to generate music was calculated by letting each model generate a certain number of music of fixed length (16 bars), and then calculating its average time to generate music. As can be seen from the table, the GCT model designed in this paper obtains the optimal results among the compared models, reducing the number of model parameters and the average time required to generate music by 0.98MB and 3.22s, respectively, compared to the Polyphony RNN model. And compared with the Pop music transformer model, the number of parameters and the average generation time were reduced by 81.69% and 55.53%, respectively. Therefore, the overall structure of the GCT model is more lightweight and has higher efficiency in generating traditional music, which provides support for the protection and inheritance of traditional music.

Table 1: Comparison of model parameters and results

Model	Characterization	Parameter quantity	Average time
GCT	MA-REMI	2.41MB	6.35s
RM-Transformer	REMI	42.78MB	29.61s
Polyphony RNN	midi-like stream	3.39MB	9.57s
Pop music transformer	MA-REMI	13.16MB	14.28s

The second comparison experiment was an objective comparison of the generated music, where each of the four models generated 40 pieces of music each, totaling 160 pieces. Taking the number of notes as an evaluation index, the statistics of the number of notes in the generated music samples can indirectly indicate whether the models can generate music with different rhythms and moods, which is a way to test the creativity of the models. Figure 3 shows the statistical results of the number of notes after generating traditional music by different models.

The statistics of the number of notes show that the average value of the number of notes obtained from the traditional music tracks generated by the GCT model (168.27) is the smallest and is generally not much, indicating that the emotions of the traditional music obtained from the generation are expressed in a slower and softer way. In contrast to the RM-Transformer model, the number of notes generated (201.25) is the highest and generally high, indicating that the emotions of the traditional music generated are more intense. While Polyphony RNN and Pop music transformer auto-composition models have higher standard deviation than the two Transformer network-based models, while generating some traditional music with more notes and some with fewer notes. To a certain extent, it shows that the traditional music generated by them is richer in emotional expression and can be better generated with different rhythms and emotions.

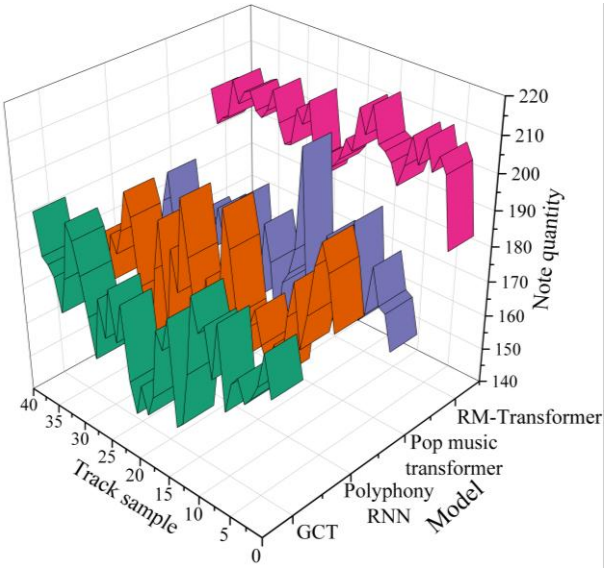

Figure 3: The result of the note count statistics

4.1.2 Different model generation results

For the traditional music generation model, if the loss function is too small due to too much training, then there may be a situation where the traditional music is not lively enough or sounds like there is no change, and such traditional music is not a good traditional music. Therefore, when training the model, it is necessary to find an appropriate stopping time that enables the model to generate the best-sounding traditional music.

Specifically, the stopping time of training is controlled by adjusting the number of iterations of the model. When the loss function reaches a certain threshold, it is sufficient to stop training and generate traditional music. The training can be stopped at different iteration times and different tunes can be generated according to the same rules, and the highest rated iteration times can be found by rating different tunes. Different tunes are selected for each iteration number to take their average score and their specific results are shown in Fig. 4.

As can be seen from the figure, with the increasing number of iterations, the traditional music scores generated by the model gradually overlap with the values of the manual scores. After the number of iterations exceeds $3.5*10^4$ times, the results of manual scoring and model scoring are relatively close. Therefore, the number of iterations for the GCT model in this paper is set to $3.5*10^4$ times to obtain the optimal traditional music generation results.

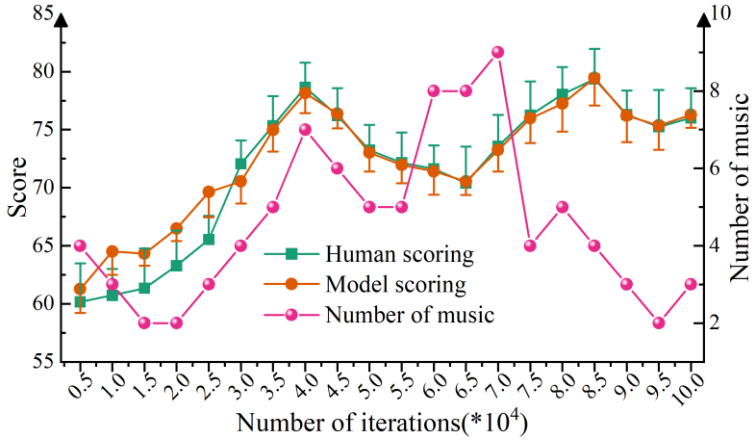

Figure 4: Music scores for different iterations

The Keras deep learning framework was used for the AI AI composition and final generation of MIDI music. A certain amount of MIDI experimental material is used in the experimental process, and the neural network is used to allow the AI to train and generate the neural network model based on the provided music material, and finally generate the composition model. In order to compare the effect of the GCT model generating traditional music and the traditional neural network model composing, this thesis uses the two models to amplify the traditional music using the same sections respectively, and scores the final generated traditional music. This time, a total of 10 sections of music were selected, and the duration of the selected sections was increased from 5 seconds to 35 seconds, and the comparison results were obtained as shown in Fig. 5. It can be seen that when the two models are performing traditional music expansion, the longer the original music section is, the higher the score of the music made. It can also be seen that the performance of the GCT model used in this paper is better than the performance of the music made using the Keras deep learning framework, and the general score is more than 12 points higher. Therefore, the use of GCT model for traditional music generation can help promote the expansion of traditional music, so that it can be renewed in the technological era with new forms of expression and inheritance paths.

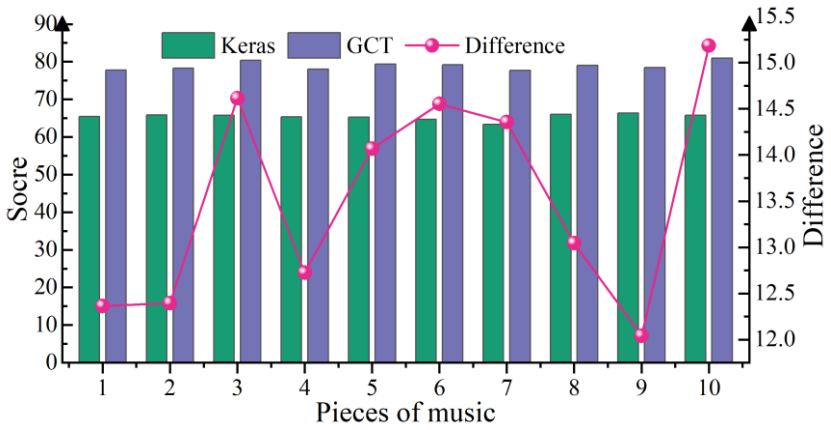

Figure 5: The generation scores of different models

4.2 Analysis of the quality of traditional music production

4.2.1 Subjective evaluation of musical style

For traditional music styles, this paper mainly categorizes them from three types: ethnic, popular and drama. For the generated music of different styles, we invited several listeners to score them, and the listeners were asked to evaluate the generated music of previous models in addition to the GCT model in this experiment.

The scoring consists of two main aspects, firstly, the listeners are invited to evaluate the pleasantness of the generated music, and the other is to evaluate the stylistic aspects of the generated music, which is a measure of whether the generated music style meets the expectations of the listeners, with a maximum of 5 points and a minimum of 1 point. A weight is added to the results of both aspects to obtain the final score. The weighting also depends on the audience, i.e. we count the number of listeners who think that melody is more important than style, and the number of listeners who think that style is more important, and the total number of people who think that melody is more important than style, and the weighting of these two aspects is the weighting of the corresponding aspect. According to the statistics, about 80% of the listeners think that the pleasantness of a piece of traditional music is more important,

while the remaining 20% pay more attention to the style of traditional music. The results of the comparison of subjective evaluation scores of different types of traditional music generated by different models are shown in Table 2.

From the data in the table, it can be seen that under different traditional music style types, the scores in terms of melody and style are significantly higher than those of the comparison models. The overall scores under the three traditional music styles of ethnic, popular and theater are 3.480, 3.442 and 3.422 respectively, which are 16.94%, 18.20% and 21.00% higher than the overall scores of the RM-Transformer model, which is the next best performer, on the three types respectively. This is due to the fact that the model in this paper mainly relies on the chord matching generator to obtain the basic chords of traditional music, and then matches them with the melody generator to obtain traditional music that is more in line with listeners' preferences. Therefore, the addition of a model with a music style extraction module and a style classifier can indeed control the generation of traditional music styles to a certain extent, and also ensure the quality of the generated traditional music.

Table 2: Subjective evaluation score comparison results

Type-		LSTM	GAN	RM-Transformer	GCT
Ethnic	Melody	2.36	2.51	3.28	3.37
	Style	1.75	1.82	1.76	3.92
	Total	2.238	2.372	2.976	3.480
Popular	Melody	2.25	2.63	2.81	3.35
	Style	1.87	1.79	3.32	3.81
	Total	2.174	2.462	2.912	3.442
Dramatic	Melody	2.35	2.64	2.74	3.36
	Style	1.57	1.72	3.18	3.67
	Total	2.194	2.456	2.828	3.422

4.2.2 Analysis of auditory subjective evaluations

Auditory subjective evaluation is the process of expressing subjective auditory feelings in words based on the objective characteristics of the audio material. This project combines the objective acoustic characteristics of traditional music, determines the musical acoustic terminology through subjective evaluation experiments and selects suitable statistical methods to count them, adopts the questionnaire survey method and expert argumentation, and finally selects 12 listeners to determine the terminology evaluation system of this traditional music evaluation. This terminology system consists of 7 items, mainly including timbre integration, acoustic balance, dynamic range, sensitivity, treble, midrange and bass, characterized as TJ1~TJ7, and the 7 items are rated on a 5-level scale of Likert scale. Figure 6 shows the results of auditory subjective evaluation of traditional music.

As can be seen from the figure, the traditional music generated using the GCT model scored above 37 points in the subjective evaluation of listening for timbre integration, acoustic balance, dynamic range, midrange and bass, while sensitivity and treble scored higher than 34.5 points. This indicates that the GCT model is more capable of extracting the timbre and acoustic characteristics of traditional music, with a more transparent midrange sound, a more cleansing bass, a greater dynamic range, and a heavier sense of musical thickness. Therefore, relying on the GCT model can significantly promote the innovative performance of traditional music and provide a new path to realize the protection and inheritance of traditional music.

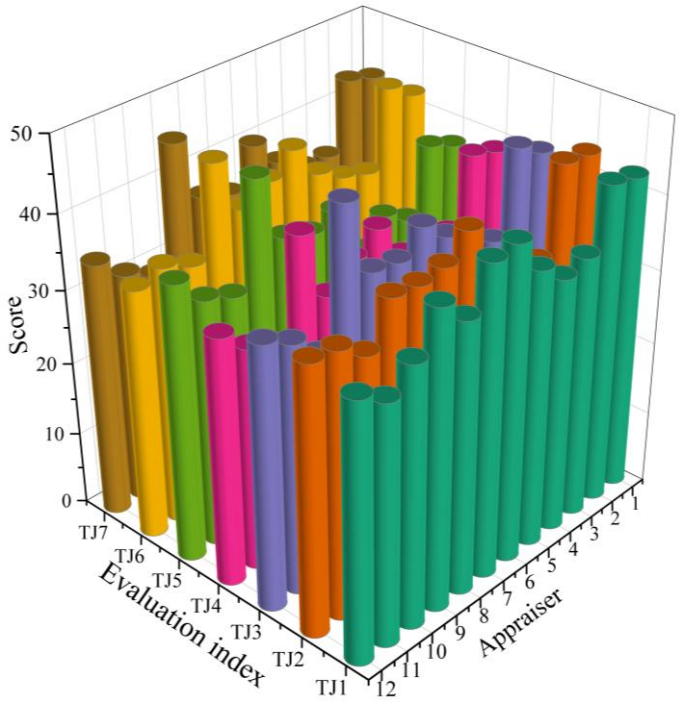

Figure 6: Subjective auditory evaluation of traditional music

5 Conclusion

Based on the establishment of the CT Music database, the article designed a traditional music generation model based on the Transformer model and validated its effectiveness by subjective and objective scores. It is found that the traditional music scores generated by the GCT model are generally higher than 12 points, and the auditory subjective scores are all higher than 34.5 points. Combining neural network technology with traditional music can significantly promote the innovative generation of traditional music and provide a new path for the protection and inheritance of traditional music.

About the Author

Jiangzhao Ye, Ph.D candidate in Jeonbuk National University. Master of Music graduated from the Zhejiang Conservatory of Music in 2023. His research interests include Musicology, Guzheng Performance, Music education, Chinese traditional musicology.

References

[1] Chow, S. M. Y. (2019). Redefining essence: tuning and temperament of Chinese traditional music. In Of essence and context: Between music and philosophy (pp. 255-267). Cham: Springer International Publishing.

[2] Peiyu, L., & Yodwised, C. (2024). ETHNOMUSICOLOGICAL METHODS IN THE TEACHING OF CHINESE TRADITIONAL MUSIC. Asia Pacific Journal of Religions and Cultures, 8(2), 643-654.

[3] Lee, B. H., & Mazzola, G. (2025). Mathematical model of modulation in traditional Chinese music: theory and computational implementation. Journal of Mathematics and Music, 19(2), 128-144.

[4] Shi, Q. (2021, September). The Study on the Development of Traditional Music in Internet Age. In Proceedings of the 5th International Conference on Algorithms, Computing and Systems (pp. 1-5).

[5] Lei, L. (2024). The latest technological developments in Chinese music education: Motifs of national musical culture and folklore in modern electronic music. Education and Information Technologies, 29(9), 10595-10610.

[6] Zhang, X. (2024). Inheritance and Innovation of Chinese Folk Song Vocal Music: Tradition and Integration in Modern Vocal Perform. Philosophy and Social Science, 1(4), 85-94.

[7] Yuan, Y. (2025). A Study on the Application of Chinese Traditional Music Elements in Modern Piano Art Creation. Lecture Notes in Education, Arts, Management and Social Science, 3(6), 270-275.

[8] Tan, Y., & Conti, L. (2019). Effects of Chinese popular music familiarity on preference for traditional Chinese music: Research and applications. Journal of Popular Music Education, 3(2), 329-358.

[9] Liu, Y., & Song, Y. (2025). The role of Chinese folk ritual music in biodiversity conservation: an ethnobiological perspective from the Lingnan region. Journal of Ethnobiology and Ethnomedicine, 21(1), 6.

[10] Zhang, J. (2025, May). Optimizing Artificial Intelligence Algorithms for Enhanced Teaching and Digital Preservation of Folk Music. In Proceedings of the 2025 2nd International Conference on Digital Society and Artificial Intelligence (pp. 185-190).

[11] Li, L. (2021). Artificial intelligence: An earthquake in the copyright protection of digital music. In Regulating Artificial Intelligence in Industry (pp. 99-113). Routledge.

[12] EZUGWU, S. I. (2025). IMPACT OF ARTIFICIAL INTELLIGENCE (AI) ON IGBO TRADITIONAL MUSIC AND CULTURE: TOWARDS FOSTERING ENTREPRENEURSHIP AMONG SECONDARY SCHOOL STUDENTS, FOR SUSTAINABLE DEVELOPMENT. AWKA JOURNAL OF RESEARCH IN MUSIC AND THE ARTS (AJRMA), 18(1).

[13] Kaliakatsos-Papakostas, M., Floros, A., & Vrahatis, M. N. (2020). Artificial intelligence methods for music generation: a review and future perspectives. Nature-inspired computation and swarm intelligence, 217-245

[14] Zheng, X., Li, D., Wang, L., Zhu, Y., Shen, L., & Gao, Y. (2017, February). Chinese folk music composition based on genetic algorithm. In 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT) (pp. 1-6). IEEE.

[15] Epstein, M. M. (2025). Artificial Intelligence and Music Mash-Ups: Monetizing an Opt-In Closed Universe Database to Preserve Royalties and Credit for Composer and Sound

Recording Rights Holders. Marquette Law Review, 108(3), 809.

[16] Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: a review. Artificial intelligence review, 55(1), 589-656.

[17] Si, Y., & Li, X. (2024, November). Overview of the Application of Artificial Intelligence in Music Creation. In Proceedings of the 2024 International Conference on Artificial Intelligence, Digital Media Technology and Interaction Design (pp. 559-565).

[18] Novelli, N., & Proksch, S. (2022). Am I (deep) blue? music-making ai and emotional awareness. Frontiers in Neurorobotics, 16, 897110.

[19] Sun, F. (2024). Analysis of virtual Reality-based music education experience and its impact on learning outcomes. Scalable Computing: Practice and Experience, 25(6), 4755-4762.

[20] Yuldashev, A. (2024). Music in Virtual Reality: New Opportunities for Composers and Performers. Jurnal Pendidikan Non formal, 1(3), 8-8.

[21] Yi, K., Wu, Y., Liu, Y., & Xu, Z. (2024). Immersive empathy in digital music listening: Ideas and sustainable paths for developing auditory experiences in museums. Sage Open, 14(2), 21582440241256339.

[22] Wei, Y. (2025). Immersive experience arousal process of vocal music language: From perspectives of “music” and “lyrics”. SAGE Open, 15(3), 21582440251356860.

[23] Wycisk, Y., Sander, K., Kopiez, R., Platz, F., Preihs, S., & Peissig, J. (2022). Wrapped into sound: development of the immersive music experience inventory (IMEI). Frontiers in Psychology, 13, 951161.