International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 4526-4535

https://doi.org/10.70517/ijhsa464296

Multimodal Intelligent Generation of Opera Styles and Musical Melodies: Time Series Analysis Strategies

Jiping Liu[1] and Mei Huang[1,*]
[1] Art College, Wanxi College, Lu'an, Anhui, 237012, China
Corresponding authors: (e-mail: hmcd--1999@163.com).

Abstract Chinese opera music, as a traditional cultural treasure, carries deep historical heritage and unique artistic charm. In this paper, a Transformer-based melody generation model for opera music, Tr-MTMG, is proposed, which realizes the inheritance and innovation of opera style through multimodal time series analysis. Methodologically, the model consists of three parts: data preprocessing network, learning network and generative network, in which the learning network contains six Encoding layer sub-networks, and the Cross-track attention mechanism is used to interactively learn the time series information between different tracks. The experimental results show that Tr-MTMG generates 128 bars of opera music with 13-19 themes, the rate of empty bars is reduced by 1.429%, the ratio of qualified notes is increased to 96.862%, and the overall quality score of subjective evaluation is 3.73 points. The model effectively solves the deficiencies of traditional music generation methods in long-term structural consistency and style maintenance, and generates opera music with rich melodic variations and good structural coherence, which provides technical support for the digital inheritance of opera music.

Index Terms Multimodal time series analysis, Opera music, Melody generation, Transformer, Cross-track attention mechanism, Style inheritance

I. Introduction

Chinese opera is a comprehensive theater art that includes various factors such as literature, music, dance, fine arts, acrobatics and performing arts [1]. Opera music is an important part of the art of opera, which includes vocal part of the singing, rhyme and instrumental part of the accompaniment, opening and passing music, it is a comprehensive art that closely combines music and drama, which has the auditory characteristics of music art, but also has to serve for the play, which needs to be closely combined with the plot, which needs to shape the characters' musical image, which needs to be closely matched with the actors' performance movements and the rhythm of the stage [2]-[5]. It should not only have the characteristics of the opera's genre, local characteristics and vernacular flavor, but also have the main vocal cadences representing the musical characteristics of the genre closely integrated with the local language (i.e., the vernacular vernacular) [6], [7]. It should have the spirit of the times (different era characteristics respond to different times), but also profoundly express the play and shape the characters, so that the audience is happy to accept, easy to circulate, after listening to the aftertaste, there is a mood [8], [9]. Its singing (singing, singing), playing (accompaniment), reading (language, recitation), playing (percussion), are required to be harmonious and complete, each with its own characteristics, its form and content is a unique style [10].

And with the development of the times, under the impact of various new cultures, opera music faces unprecedented challenges. In this context, the role of artificial intelligence in the inheritance and protection of opera music is becoming more and more important, artificial intelligence can be based on the traditional opera music, generate the corresponding melody, provide technical support for the development of opera music in the new era, and is conducive to the inheritance of opera music style [11]-[14].

Based on the artistic characteristics of opera music, this study explores the innovative path of combining advanced deep learning technology with traditional music theory. Starting from the multi-track characteristics of opera music, the study considers different instrumental tracks as time sequences of different modalities, and constructs a generative model capable of learning the deep features of opera music by analyzing the interaction and temporal dependence between the tracks. In terms of technology, the Transformer architecture is chosen as the basic framework, and the Cross-track attention mechanism is innovatively designed to realize the interactive learning of information between multi-tracks. By constructing a complete technical system containing data preprocessing, feature learning and music generation, we strive to maintain the traditional style of opera music while giving it new vitality, and provide intelligent solutions for the inheritance and development of opera music.

II. Transformer-based melody generation model for opera music

Intelligent music generation has always been one of the hottest research directions in music-related fields. In this paper, in order to better inherit the opera style, we innovatively construct a Transformer-based opera music melody generation model Tr-MTMG, which is capable of generating multi-track music melody in opera style, and it consists of three parts: data preprocessing network, learning network and generation network.

II. A. Data processing and representation

In this paper, we collect a large amount of opera music to construct a dataset that meets the experimental requirements, which contains 500 pieces of multi-track music in MIDI format containing labeled instrumental tracks, and we use Pretty-midi to filter the dataset according to the labeled instrumental tracks, retaining the musical melodies in 4/4 time and filtering music with large differences in duration and song patterns.

II. B. Transformer

Transformer is a parallel training model from sequence to sequence, with a structure divided into four main parts: input (positional encoding), encoder, decoder, and output.

Since the Transformer model itself does not have the inherent sequential ability to handle sequence data like an RNN, a positional encoding mechanism is introduced. The encoding method used in Transformer is sine-cosine encoding, and the formulas for static sine-cosine encoding are shown in equations (1) and (2):

$$PE\left(pos,2i\right)=\sin\left(\frac{pos}{10000^{\frac{2i}{d_{\mathrm{mod}\,et}}}}\right) \tag{1}$$

$$PE\left(pos,2i+1\right)=\cos\left(\frac{pos}{10000^{\frac{2i}{d_{moki}}}}\right) \tag{2}$$

where pos denotes the absolute position of the object, d_{model} denotes the learning object dimension, and i denotes the dimension index value.

II. B. 1) Self-attention mechanisms

The attentional mechanism allows each position of an input sequence to be able to attend to all other positions in the sequence, thus learning the relationships between positions. The computation of the self-attention mechanism can be realized by means of matrix operations that allow the attention weights of all positions in the whole sequence to be computed at the same time. It is divided into the following steps:

(1) Linear Transformation

Linear transformation is performed on the representation of each position in the input sequence to map it into a new representation space, generating three key variables, which are query key Q, key value key K, and value key V.

(2) Attention Score Martix

By performing dot product operation on the transformed Q, K, and V representations, the Attention Score Matrix is obtained, each element of the matrix represents the attention score between two sequence positions, and the parameters are calculated as shown in equation (3):

$$Attention(Q,K,V)=Softmax\left(\frac{QK^{t}}{\sqrt{d_k}}\right)V \tag{3}$$

where Q,K and V represent query vectors, key vectors and value vectors, respectively, and d_k represents the dimensions of query vectors and key vectors.

(3) Attention Score Normalization

Apply Softmax function to each row of the Attention Score matrix, Softmax function can convert a multidimensional vector z into a vector $\sigma(z)$ of the same dimension as shown in equation (4):

$$\sigma(z_i)=\frac{e^{z_i}}{\sum_{i}e^{z_i}} \tag{4}$$

where z is the input vector, z_i is the i th element in the vector z, and $\sigma(z_i)$ is the i th element in the vector after the Softmax function.

(4) Weighted Sum

The final attentional output representation is generated by applying the normalized attentional weight matrix to the original input sequence for weighted combination. To increase the expressiveness and learning ability of the model, the Transformer model employs Multi-Head Self-Attention, which captures different relationships by computing multiple sets of attention weights in parallel.

Final Representation Mapping: the spliced Multi-Head Attention representation is mapped to the desired representation space by another linear transformation, as shown in Eqs. (5) and (6):

$$MultiHead\left(Q,K,V\right)=Concat\left(Head_1,\cdots,Head_n\right)W^O \tag{5}$$

$$Head_i=Attention\left(QW_i^Q,KW_i^K,VW_i^V\right) \tag{6}$$

where Q,K,V represent the query matrix, key matrix, and value matrix, respectively. W_i^Q,W_i^K,W_i^V represent the corresponding weight matrices when computing Q,K,V, respectively. W^O denotes the weight matrix of the original matrix dimension after re-projecting the spliced multi-head attention matrix. i denotes the index of different Heads, d_k is the dimension of the query key, and Concat is the matrix splicing operation.

II. B. 2) Feedforward Neural Networks

Feedforward neural network (FFN) is one of the most basic neural network structures consisting of multiple layers of neurons with full connectivity between the layers. In the Transformer model, feedforward neural networks are used to perform nonlinear transformations in each sublayer of the encoder and decoder.

(1) Structure

A feedforward neural network consists of multiple layers of subneurons, with the previous and subsequent layers fully connected in each layer. Typically, a feedforward neural network contains at least one hidden layer and one output layer, where there can be more than one hidden layer.

(2) Activation function

In each neuron of a feedforward neural network, an activation function is introduced to introduce a nonlinear transformation.

Rectified Linear Unit: when the output is greater than 0, the output is equal to the input, and when the output is less than or equal to 0, the output is 0, as shown in equation (7):

$$f(x)=\max(0,x) \tag{7}$$

Sigmoid: maps the input to between 0 and 1 as shown in equation (8):

$$\sigma(x)=\frac{1}{1+e^{-x}} \tag{8}$$

where x denotes the input and $\sigma(x)$ denotes the output of the Sigmoid function.

Tanh (hyperbolic tangent function): maps the input to between -1 and 1, and the output is symmetric around 0, as shown in equation (9):

$$\tanh(x)=\frac{e^x-e^{-x}}{e^x+e^{-x}} \tag{9}$$

where x denotes the input and $\tanh(x)$ denotes the output of the Tanh function.

(3) Parameter learning

The parameters of the feedforward neural network include the weights and biases of each layer. These parameters need to be learned by the back-propagation algorithm and optimizer, enabling the neural network to fit the training data and achieve good generalization performance on unseen data.

(4) Forward propagation

In the forward propagation process, the input data is passed layer by layer through the linear transformation and excitation function of each layer to finally get the output as shown in equation (10):

$$FFN(x)=\max\left(0,xW_1+b_1\right)W_2+b_2 \tag{10}$$

where x represents the input data, W_1 represents the weight between the input layer and the output layer, b_1 represents the bias term of the hidden layer, W_2 represents the weight between the hidden layer and the output layer, b_2 represents the bias term of the output layer, and $\max(0,x)$ represents the activation function.

(5) Backpropagation

The process of backpropagation is to propagate the error from the output layer back to the hidden and input layers, updating the weights and biases of each layer to make the output of the network closer to the true label.

II. C.Network Modeling

The Tr-MTMG model mainly improves on Transformer with a learning network that can be used to learn time series information between different modal tracks, and generates operatic music pieces through a generative network based on the learned time series information.

II. C. 1) Learning networks

The learning network contains 6 Encoding layer sub-networks, and each Encoding layer sub-network can learn two-by-two interactions with audio tracks from real samples. Each layer subnetwork structure contains six Encoding modules, and the Cross-track attention mechanism is an important part of the Encoding modules.

Cross-track attention mechanism is mainly improved on the basis of Self-attention mechanism, which is different from Self-attention mechanism in that Self-attention mechanism mainly learns the information of the sequence itself, while Cross-track attention mechanism is different from the Self-attention mechanism, which mainly learns the information of the sequence itself, while the Cross-track attention mechanism learns the information of different sequences. Therefore, in the case of learning time series information between different modal tracks, this paper selects the Cross-track attention mechanism for learning between tracks. The opera music track is treated as the learning sequence and the generative music track is treated as the learned sequence, which are denoted by $X_p \in R^{T_p \times d_p}$ and $X_g \in R^{T_i \times d_s}$, respectively. $T_{(\cdot)}$ is denoted as the length of the sequence and $d_{(\cdot)}$ denotes the dimension of the feature.

In the sequence learning process, the Cross-track attention mechanism treats the query as the dot product of the inputs of the learned target and the input transformation matrix, i.e., $Q_p = X_p W_{Q_p}$, and the key-value pairs as the dot product of the inputs of the learned target and the key-value pair transformation matrix, i.e., $K_g = X_g W_{K_g}$ and $V_g = X_g W_{V_g}$, where $W_{Qe} \in R^{d,xd_e}$ is the weight of the transformation matrix for the query of the input sequence of the opera music and $W_{\kappa_x} \in R^{d_x \times d_x}$ and $W_{v_x} \in R^{d_x \times d_x}$ are the transformation matrices for generating the key-value pairs of the musical input sequences respectively The weights, i.e., the Cross-track attention value from the opera music sequence track to learning the information of the generated music sequence can be expressed as Equation (11):

$$Z_{p\to g} = CT_{p\to g} attention(X_p, X_g) = soft\max\left(\frac{Q_p K_g^T}{\sqrt{d_k}}\right) V_g \tag{11}$$

Among them, $Z_{p\to R}$ is one of the multi-head Cross-track attention mechanisms, as shown in Equation (13), which can be regarded as $head_h$. Next, the h Cross-track attention values are spliced, and then linear activation is applied to them, as shown in Equation (12), to obtain the Multihead Cross-track attention value $Multihead(h)$:

$$Multihead(h) = W[head_1; head_2; ...; head_h] \tag{12}$$

$$head_h = CTattention(Q_p, K_g, V_g) \tag{13}$$

The three tracks need six Encoding layers to learn from each other, and each layer contains six Encoding modules. When the opera music sequence is the source sequence and the generated music sequence is the target sequence, the model needs to have two Encoding layer sub-networks, and one layer is used for the opera music sequence to learn the melodic information of the generated music sequence as in Eq. (14). One layer is used for the opera music sequence to learn the rhythmic information of the generated music sequence, as in Equation (15). Where $\hat{Z}_{p\to s}^{[i]}$ is the sequence output after generating the music sequence by opera music sequence learning, after passing through the Cross-track attention mechanism of the i-layer polytope, $i = \{1,2,3,4,5,6\}$:

$$\hat{Z}_{p\to g}^{[i]} = CT_{p\to g}^{[i]}(LN(Z_{p\to g}^{[i-1]}), LN(Z_p^{[0]})) + LN(Z_{p\to g}^{[i-1]}) \tag{14}$$

$$\hat{Z}^{[i]}_{p\to b} = CT^{[i]}_{p\to b}(LN(Z^{[i-1]}_{p\to b}), LN(Z^{[0]}_{p})) + LN(Z^{[i-1]}_{p\to b}) \tag{15}$$

After obtaining the value of the multi-head Cross-track attention sequence by Equation (12), the output sequence is then allowed to go through layer normalization to obtain the output sequence with the same dimension as the input sequence, and the obtained output sequence is used as the input of the feed-forward sub-layer, and after the feed-forward sub-layer, the output sequence with the same dimension as the input sequence is then connected by residuals to obtain the sequences $Z^{[i]}_{p\to g}$ and $Z^{[i]}_{p\to b}$, which are the output sequences after encoding in i layers. They are Eq. (16) and Eq. (17), respectively:

$$Z^{[i]}_{p\to g} = f_{\theta^{[i]}_{p\to g}}(LN(\hat{Z}^{[i]}_{p\to g})) + LN(\hat{Z}^{[i]}_{p\to g}) \tag{16}$$

$$Z^{[i]}_{p\to b} = f_{\theta^{[i]}_{p\to b}}(LN(\hat{Z}^{[i]}_{p\to b})) + LN(\hat{Z}^{[i]}_{p\to b}) \tag{17}$$

After obtaining two output sequences $Z^{[l]}_{p\to g}$ and $Z^{[i]}_{p\to b}$ for learning the other tracks, they are spliced as in Eq. (18), and the final output is the sequence of opera music Z_p:

$$Z_p = Concat(Z^{[i]}_{p\to g}, Z^{[i]}_{p\to b}) \tag{18}$$

II. C. 2) Generating networks

After learning the information between the tracks, the generative network will generate the opera music based on the time series information learned in the learning network. The Transformer model, i.e., GPT model, after removing the Self-attention mechanism and the encoder module, is utilized as the generative network. It consists of an embedding layer, six decoder modules and a linear Softmax layer, each decoder module consists of eight 256-dimensional Self-attention layers and 1024-dimensional feedforward sublayers, which are capable of predicting the next momentary state based on the previous state.

II. C. 3) Model training

The loss function is an important part of the deep learning model, which can guide the network how to get the sequence that meets people's needs, so the construction of the loss function is especially important. In order for the model to be more effective in supervised learning, the Teacher Forcing method is chosen for training in this paper, i.e., regardless of whether the predicted result is the same as the real structure or not, it takes the notes of the real data as the input, so as to predict the notes of the next moment:

$$L(\theta) = -[y_t \log y_t + (1 - y_t)\log(1 - y_t^{'})] \tag{19}$$

y_i is the real state at a certain moment, and $y_i^{'}$ denotes the state obtained from the prediction at that moment. $L(\theta)$ denotes the cross-entropy loss function, and the learning rate is set to 0.0001. The dataset is 3,000 MIDI files with 4/4 beats labeled with musical instruments, and 2,000 MIDI files in the dataset are classified as the training set and 1,000 MIDI files are classified as the testing set. During model training, the threshold of the number of training iterations set in this paper is 10000.

III. Analysis of the results of the evaluation of the experiment

The Tr-MTMG opera music melody generation model based on multimodal time series adopts a combination of subjective evaluation and objective indicators, which is more comprehensive, objective and reasonable. This evaluation method can provide accurate and reliable feedback for the improvement of the model to ensure the reliability and stability of the model.

III. A. Objective evaluation

III. A. 1) Contrasting models

In this study, we use homemade datasets to conduct relevant experiments, and select Music transformer and Theme Transformer, the two most typical open source models for music generation, as comparison models, which will be introduced in the following respectively.

(1) Music Transformer: a music generation model based on relative attention mechanism, which can generate music with high long-term consistency.

(2) Theme Transformer: a Transformer-based model that proposes a novel positional encoding method and a method for balancing the attention mechanism, specifically for generating music with themes.

III. A. 2) Evaluation indicators

There is no common standard for objectively evaluating the quality of computerized music generation. Therefore, there are relatively few existing objective evaluation methods. The multi-track sequence generation adversarial network model proposes some objective evaluation metrics based on some features of music data, including:

Number of themes (SQ): used to evaluate the ability of the model to generate music themes, the more the number of themes contained in the music generated by the model under the same bars, the stronger the ability of the model to generate music themes.

Empty bar rate (EB): the ratio of the number of bars without notes in the track to the total number of bars in the music.

Number of Used Note Phoneme Categories (UPC): the number of different phoneme names contained in each measure of a music sample, ranging from 0 to 12.

Qualified Note Ratio (QN): is the ratio of the number of measure notes in each measure of the generated music sample that are qualified notes. Qualified notes in this context refer to notes whose duration is not less than three standard time steps (32-cent notes), otherwise they are considered as unqualified notes. This metric reflects whether the notes of a music sample are too spread out.

III. A. 3) Number of topics generated

Theme fragments in an opera music are often used to express the tone and feelings of the music, and will be repeated in a song, which is the core of a song. In this paper, the number of theme fragments in the generated opera music is counted to visualize the model's theme generation ability and performance level in music composition. The higher the number of themes contained in the opera music generated by the model, the stronger the model's ability in theme generation in opera music. On the contrary, a smaller number of themes contained in the opera music generated by the model implies that the model performs poorly in generating themes for opera music.

In order to explore the ability of each model in generating theme fragments when composing music of different lengths, in this paper, Music Transformer, Theme Transformer, and Tr-MTMG were allowed to generate 60 pieces of opera music of 32, 64, and 128 bars, respectively, and record the number of theme fragments for each piece of opera music. The results of the number of themes of the generated opera music are shown in Fig. 1, Fig. 2 and Fig. 3, where the horizontal coordinates indicate the different models, the vertical coordinates indicate the number of themes of the opera music generated by each model, and the image portion in the coordinate area indicates the density distribution of the number of themes.

As the number of music bars doubles, the number of opera music themes generated by all models does not show a corresponding multiplication trend, which indicates that the ability of all models to generate long-term structured music is inferior to the ability to generate short-term structured music. In addition, a side-by-side comparison of Tr-MTMG with other models reveals that the distribution of the number of themes in the music generated by Tr-MTMG is relatively higher regardless of the number of music bars, with the number of themes ranging from [3,7], [6,12], [13,19] for generating 32, 64, and 128 bars of operatic music, which suggests that the Tr-MTMG model has a stronger ability to generate themes for opera music.

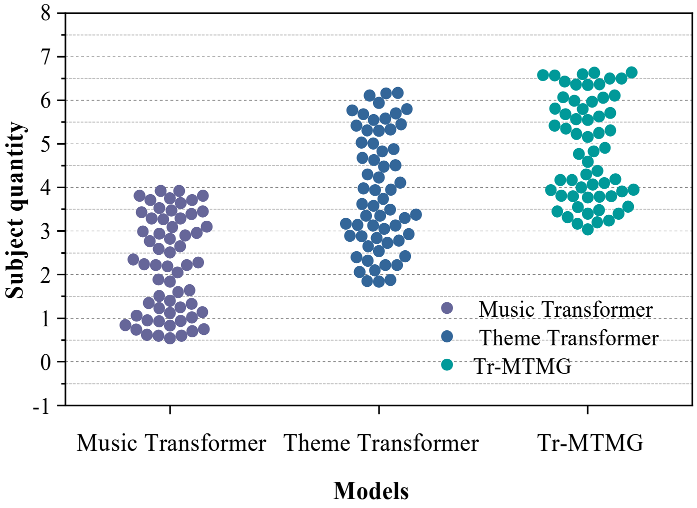

Figure 1: The number of music topics in 32-bar topics

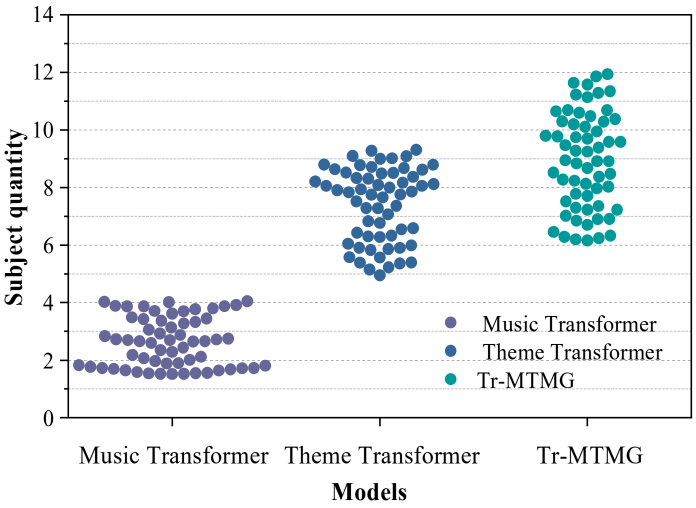

Figure 2: The number of music topics in 64-bar topics

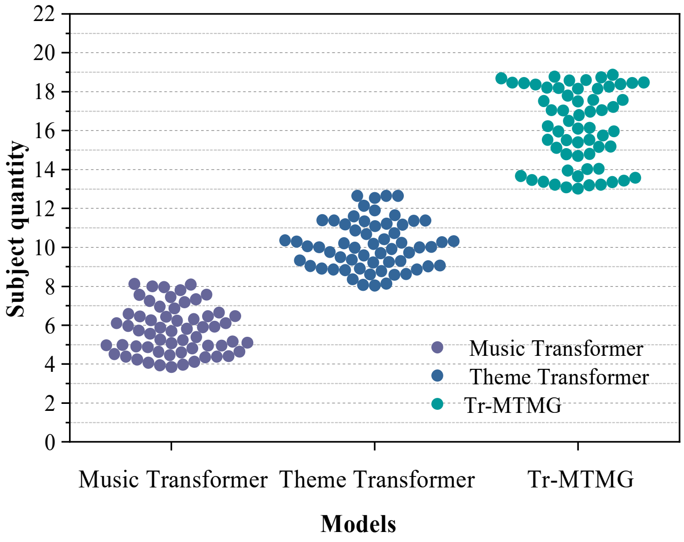

Figure 3: The number of music topics in 128-bar topics

III. A. 4) Evaluation of indicator results

In order to make the comparison results more reliable and fair, this paper uses the same objective evaluation metrics to initially evaluate the music samples generated by the 3 models to reflect the performance merits of the models. In this paper, the 3 models were trained on the same dataset and the samples generated by them were quantized. To derive the objective assessment metric scores, the PyPianoroll tool was used to process the resultant samples.

A comparison of the objective metrics results is shown in Table 1. In the empty bar rate (EB) metric, the generated opera music has a different degree of reduction compared to both Music transformer and Theme Transformer models, and the EB values of the opera music have been reduced by 0.732% and 1.429%, respectively, and the reduction of empty bar rate can make the phrases more coherent and natural, and also make the structure of the whole piece more obvious and logical.

The Tr-MTMG model is significantly improved compared with the comparison model in terms of the number of phonetic categories (UPC), and the Tr-MTMG model generates opera music samples with richer phonetic categories, with a UPC value of 3.835, which proves that the generative model is able to produce more melodically rich opera music. The enhancement of the variety of sound names can bring richer melodies to the music and increase more possibilities of generating music samples.

In the Qualified Note Ratio (QN) index, the Tr-MTMG model has significant improvement compared with the comparison model, and the QN index values of opera music are improved by 9.021% and 3.777% respectively, and the improvement of the QN ratio can effectively improve the incoherence problem in the music.

Table 1: Comparison result of objective results

Index	Models	Violin	Piano	Ensemble
EB (%)	Music Transformer	19.505	21.147	19.149
	Theme Transformer	22.986	20.725	19.846
	Tr-MTMG	18.674	18.828	18.417
UPC	Music Transformer	2.124	1.646	3.239
	Theme Transformer	2.643	1.718	2.805
	Tr-MTMG	3.808	2.355	3.835
QN (%)	Music Transformer	82.422	85.212	87.841
	Theme Transformer	83.903	90.329	93.085
	Tr-MTMG	85.786	91.974	96.862

III. B. Subjective evaluation

III. B. 1) Evaluation indicators

As a product of artistic creation, music still needs human participation in its evaluation, because it is impossible to judge a work of art only with quantitative hard indicators, and only human subjective evaluation is the most convincing, so this paper designs relevant subjective evaluation indicators to evaluate the generated opera music more comprehensively, including: (1) evaluation value: whether the emotion is positive or negative, (2) arousal: whether the emotion is calm or excited, (3) authenticity: the degree of similarity with human creation, (4) harmony: the degree of melodic fluency and harmony (5) Overall quality: the overall quality level of the music.

III. B. 2) Experimental Procedures

Before the experiment began, suitable experimenters needed to be selected, and for all experimental participants their basic information including name, age, gender, and musical experience was required. The musical experience was categorized into five levels, and 30 participants, including 15 males and 15 females, were carefully selected to subjectively assess the generated opera music clips. The average age of the participants was 26 years old, and the average musical experience was 2.63 Finally, each participant was provided with 15 music clips containing 5 different emotions, with 3 music clips for each type of emotion. For the model-generated music snippets, an audio converter was used to convert the MIDI music into MP3 format for easy listening, and the music snippets were kept to 30 s or less in order to avoid the potential impact of inconsistencies in the length of the generated music on participants' ratings. Participants rated the music pieces on the five proposed metrics, with ratings increasing from 1 to 5.

III. B. 3) Evaluation results

The results of subjective evaluation scores of opera music generated by different models are shown in Fig. 4.The Tr-MTMG model is better than the other models in subjective listening experiments.The overall quality scores of opera music generated by the three models are 3.37, 3.52 and 3.73.The subjective scores of the Tr-MTMG model are 10.68% and 5.97% higher than those of the other models.The generated opera music is more with authenticity and harmony, and better in overall quality, which is conducive to the stylistic inheritance of opera music. And in terms of emotional expression, when the provided emotion is positive or negative, and the mood is excited or calm (Valence-High/low, Arousal-High/Low), the model is able to generate specific emotional music according to the provided emotional conditions, which indicates that the model is able to keenly perceive the change of the emotional conditions, and fully learns the music emotional characteristics, thus, the proposed model performs better on the emotional music generation task.

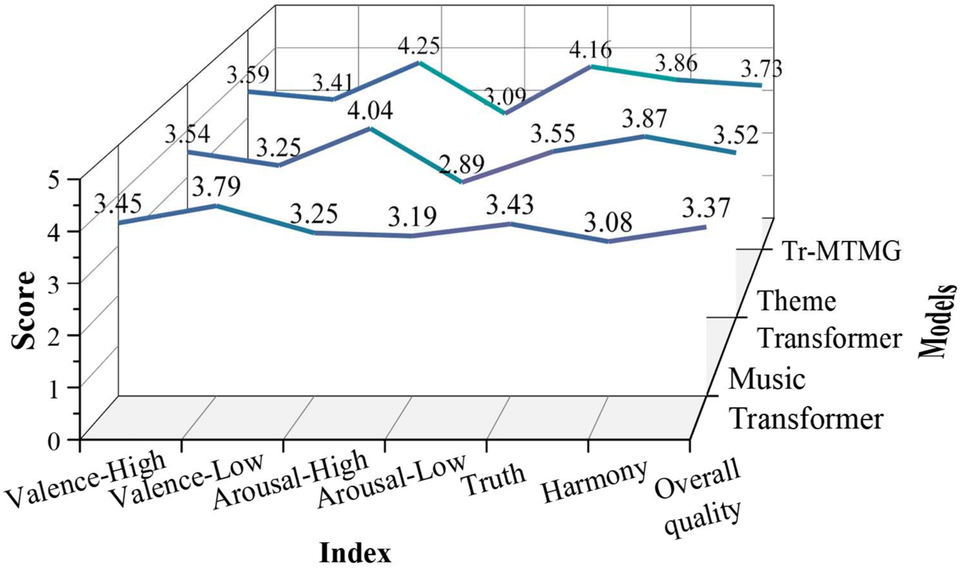

Figure 4: The subjective evaluation results of opera music

IV. Conclusion

The Tr-MTMG model in this study successfully realizes the intelligent generation and style inheritance of opera music. Objective evaluation indexes show that the model excels in musical structural integrity, and the rate of empty bars is only 18.417%, which is significantly lower than the comparison model. The tone name variety index reaches 3.835, indicating that the generated opera music has rich melodic variations. The ratio of qualified notes was as high as 96.862%, which effectively improved the coherence problem of the music. In terms of subjective evaluation, the overall quality score of the model-generated music by 30 participants is 3.73, which is highly recognized in both authenticity and harmony dimensions. The introduction of the Cross-track attention mechanism enables the model to effectively capture the interactions between multi-tracks, and the generated opera music not only maintains the traditional flavor, but also demonstrates good innovativeness. The experimental results demonstrate the feasibility and effectiveness of the method of combining deep learning techniques with opera music features. The model provides a new technical means for the digital protection of opera music, which is of great significance for promoting the innovative development of traditional culture. In the future, the model architecture can be further optimized to improve the quality of long-time music generation and be extended to the music generation tasks of more opera genres.

Funding

1. Major Scientific Research Projects of Higher Education Institutions in Anhui Province: Research on the Era Characteristics and Inheritance of the Dabie Mountain Spirit in Wanxi Lu Opera (2023AH040341).

2.Anhui Province Social Science Innovation and Development Research Project: Research on the Historical Changes and Inheritance of the Era Cognition and Presentation of Lu Opera on the West Road (2022CX149)

3.Anhui Provincial Education Science Research Project Special Project: Research on Quality Monitoring of Music Curriculum Implementation in Compulsory Education Stage in Lu'an Region (JKT24145)

4. Key Scientific Research Project of Universities in Anhui Province: Research on the Aesthetic Thoughts in the Qing Dynasty Folk Music "Liaozhai Folk Qu" (2024AH053470).

References

[1] Miller, T. E., Church, M., Reynolds, D., DeVeaux, S., Hewett, I., Hughes, D., & Katz, J. (2015). Chinese opera. The Other Classical Musics, 126-37.

[2] Rao, N. Y. (2017). Chinese Opera Percussion from Model Opera to Tan Dun. China and the West: Music, representation, and reception, 163-185.

[3] Wang, M. (2023). Analysis of Laiwu Bangzi opera in Shandong province as a resource for teaching Chinese opera music history: Teaching Chinese opera music history. International Journal of Curriculum and Instruction, 15(3), 1399-1413.

[4] Xiang, J. (2023). An analysis of the essence of Chinese opera and vocal music from the perspective of hermeneutics and reception aesthetics. Trans/Form/Ação, 47(3), e0240029.

[5] Zhang, Z. (2023). “Model opera” of the 20th century in Chinese musical culture. Notes on Art Criticism, 1(23), 206-210.

[6] Li, K. (2022). THE INFLUENCE OF CHINESE NATIONAL OPERA ON THE DEVELOPMENT OF CONTEMPORARY CHINESE VOCAL MUSIC UNDER MUSIC ANTHROPOLOGY. Psychiatria Danubina, 34(suppl 4), 770-770.

[7] Geng, Y. (2024). The application and influence of Western opera elements in Chinese opera in the 20th century. Învăţământ, Cercetare, Creaţie, 10(1), 128-137.

[8] Wu, H., Loo, C. F., & Chan, J. C. (2022). Visual Analysis of The Research Hotspots, Frontiers and Trends of Chinese Opera From 2011 To 2020. Asian Journal of Arts, Culture and Tourism, 4(4), 7-22.

[9] Wu, J., Jiang, K., & Yuan, C. (2019). Determinants of demand for traditional Chinese opera. Empirical Economics, 57, 2129-2148.

[10] Xue, Y. (2023). Analysis of Musical Performance and Characterization in Chinese Opera. Frontiers in Art Research, 5(16).

[11] Yao, M., & Liu, J. (2024). The analysis of Chinese and Japanese traditional opera tunes with artificial intelligence technology based on deep learning. IEEE Access, 12, 21084-21091.

[12] Bao, F., & Li, X. (2024). Analysis of Cross linguistic Non Material Opera Culture Communication in the Era of Artificial Intelligence. Revista Ibérica de Sistemas e Tecnologias de Informação, (E72), 607-618.

[13] Guo, S., Zhang, Y., & Sun, Z. (2024, June). Digital Empowerment of Excellent Traditional Chinese Music Culture Education. In International Conference on Human-Computer Interaction (pp. 231-241). Cham: Springer Nature Switzerland.

[14] Fang, X., Liu, C., & Yu, C. (2022). Application Research of Digital Technology in Inheritance and Development of Jiangxi Local Opera: Taking Gannan Tea Picking as an Example. In Proceedings of the 1st International Conference on Public Management, Digital Economy and internet Technology. https://doi. org/10.5220/0011736700003607.