ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Exploration of the Path of Absorption and Re-creation of Korean Traditional Music to the Music Elements of the Central Plains in the Perspective of the History of Musical Exchanges

Zhentang Cheng[1,*]

[1] School of History, School of Music, Anhui Normal University, Wuhu, Anhui, 241000, China

SUMMARY: *In this paper, MIDI data of Korean traditional music and music from the Central Plains were collected, melodic features and note features of the music data were extracted, and a multi-style melody and arrangement generation (MSMICA) model was constructed. The Multi-Sequence Generative Adversarial Network serves as the main framework of the model, including two core parts, the generator and the discriminator. The generator incorporates a reinforcement learning mechanism to reward with feedback from the discriminator, while the discriminator utilizes the GRU component for the recognition of mesogenic musical elements as well as harmonic music generation, and proves the feasibility of the model based on relevant experiments. During the training process, the Loss and Accuracy values of the model converge to about 0.44 and 0.95 successively in about 60 rounds of iterations, showing good training results. The fluctuation amplitude of the signal features of the model-generated traditional music of the Central Plains style Korean music is between ±0.8, and the signal features in the Meier spectrum, spectral roll-off and chromatic frequency are similar to the real music, and the model-generated music clip is more in line with the traditional music of the Central Plains style Korean music.*

KEYWORDS: *MSMICA model; Multiple Sequence Generative Adversarial Network; Reinforcement Learning; Korean traditional music; Central Plains music*

1 Introduction

It is recorded in the Records of the Joseon Dynasty - Taejong's Records that: "In the previous dynasty, King Gwang sent an envoy to ask for Tang musical instruments and laborers, and his children and grandchildren kept their profession, and in the Chungnyeol Dynasty, Jin Luying took charge of it, and in the Chungsusu Dynasty, his grandchildren Yu took charge of it. And according to the Song music book, Yuanfeng years, Goryeo seek musicians and teach. However, my Oriental music, really out of China. Passed down for a long time, I am afraid that there may be errors, begging and habitual supervisors to review in detail, looking for its old score, after the Tang and Song of the remains of the sound, set the Sheng Dynasty of the right music." The record shows that China and North Korea have had close exchanges since ancient times. Goryeo Wenzong reign of Korea and the Song dynasty is more frequent, political, cultural and economic exchanges are very active, a large number of rituals and music from the Song dynasty to Goryeo, Goryeo even obtained the "small China" reputation [1, 2].

The exchanges between China and Korea made the music exchanges more and more frequent. Because of political reasons, most of the traditional Korean music is dark in color, has

*chengzhentang1129@163.com
https://doi.org/10.65102/is2026071 Published: 30 April, 2026

a unified style, and has its own unique structural characteristics and attributes in musical elements such as modes, tunes, and rhythms and beats [3-5]. Facing the internationalization and opening up era in the 21st century, Korean traditional music will face a serious existential crisis and challenges, and the traditional music passed down in the folklore will be significantly reduced or declined in terms of "quantity" and "quality" [6, 7]. Therefore, it is necessary to realize horizontal development in addition to vertical development, and to draw fresh nutrients through cross-cultural, cross-ethnic, and cross-regional communication and exchange in order to make Korean traditional music always maintain a strong vitality. The music of the Central Plains has a long history, is rich in genres, and is characterized by rootedness, originality, radiation and inclusiveness, making it the most representative ethnic folk music culture [8-10]. Therefore, the musical elements of the Middle Kingdom have also become an important source for the utilization and creation of Korean traditional music.

The article extracts and analyzes the melodic and note features based on the data of Korean traditional music and Middle Chinese music, and proposes the Multi-style Melody and Arrangement Generation (MSMICA) model for the problem of integrating the elements of Korean traditional music and Middle Chinese music. The model takes the obtained melodic and note features as input, and controls the style and harmonic coordination of the generated music through a style discriminator as well as a harmony discriminator model. The MSMICA adversarial network is then used to optimize the multi-sequence co-training by sharing feedback rewards, and then generate Korean traditional music with elements of Central Plains music. Based on the collected music data, sufficient validation experiments were conducted to evaluate the effectiveness of the model.

2 Relevant data collection and processing

In order to be able to incorporate elements of Chungwon music into Korean traditional music, the study obtained a total of 422 MIDI music files of Korean traditional music styles and 347 MIDI music files of Chungwon music styles from the Internet. After the main theme extraction, the qualified MIDI main theme music files were obtained: 395 pieces of Korean traditional music and 328 pieces of Chungwon music. All the MIDI music files were in 4/4 beat, with a beat rate of 120 beats per minute, and a total of 37 types of notes were found.

2.1 Melodic feature extraction

One of the success factors for the recreation of Korean traditional music that incorporates elements from the Middle Kingdom is the generation of a melodious melody. The study will extract the entire main melody of the training samples, and in the database, in addition to the dataset with the full song, the main melody will be added to enhance the melodic part of the training.

The study uses a similarity matrix to extract the main melody. In MIDI format, if music is scanned at 10 frames per second, i.e., 1 frame per 0.1 second. Each frame of music can be described by a vector $v=\left[x_1,x_2,\cdots,x_n\right]$, with n denoting the number of notes contained inside a frame of music, and a piece of music consists of multiple frames of music, so it can be described by the matrix $X=\left[v_1,v_2,\cdots,v_N\right]^T$. where it is assumed that each piece of music has N frames. This results in a similarity matrix for a piece of music as a $(N-1)\times(N-1)$ matrix, and the distance between two pieces of music can be calculated using the Euclidean distance:

$$S_{i,j} = \sqrt{\left(v_i - v_{i+1}\right)^2 + \left(v_j - v_{j+1}\right)^2} \tag{1}$$

where $i, j = 1, 2, \cdots, N-1$.

The computation of the similarity matrix is good for finding repeated segments with small variations, which can be retrieved using the following formula:

$$S = 1 - \frac{\left\| X_i - X_{i+1} \right\|}{\sqrt{12}} \tag{2}$$

where 12 is the twelve mean law and X_i and X_{i+1} are the matrices corresponding to the two pieces of music. If the computation results in a high value of S, the two pieces of music are similar. The similar segments are stored in the trained music database using Python for that part of the melody extraction.

2.2 Note feature vector construction

After melodic feature extraction, note feature vector construction is also needed to realize the natural absorption of traditional Korean music into the elements of Central Plains music to create music with closer relevance. The whole process of note feature vector construction can be summarized in the following steps:

(1) From the extracted MIDI main melody file, the main melody note sequences are sampled and acquired according to the quantization time step, which is 0.1 seconds.

(2) Set the context window as 6, form the binary note data set of (center note, target note) as a data set, and make the binary data set as a uniquely hot coded representation with a uniquely hot coded dimension of 125.

(3) The feedforward neural network model is constructed, and its structure is composed of input layer, hidden layer, softmax layer, and output layer in a fully connected manner.

(4) The input layer inputs the center note in the binary note data set, passes through the hidden layer and the softmax layer to reach the output layer, and the output of the output layer calculates the error with the target note of the binary note data set.

(5) Backpropagate the error and iteratively train and learn until the model converges when the note feature vector to be generated is obtained from the parameters of the hidden layer, and a note has a unique note feature vector corresponding to it.

The cosine similarity function is used to accurately compute the contextual semantic information between the generated note feature vectors, the greater the similarity indicates the greater the probability that the notes will appear in each other's context in the dataset. Define two note feature vectors as X and Y:

$$\cos\left\langle X, Y \right\rangle = \frac{X \cdot Y}{\left| X \right| \times \left| Y \right|} \tag{3}$$

Use the cosine value to measure the similarity of two vectors; when they are identical, the cosine value is 1 and they are most similar. When they are orthogonal, the cosine value is 0 and they are independent.

3 Generation of Korean traditional music based on the musical elements of the Central Plains

A multi-style melody and arrangement generation model (MSMICA) is proposed to address the problem of incorporating elements of Middle Eastern music into Korean traditional music. Specifically, the multi-style arrangement generation problem and the details of MSMICA are presented.

3.1 Formal definition of the problem

Assuming a musical melody $M=\left\{m_1,m_2,\cdots,m_{l_m}\right\}$, its corresponding rhythm $R=\left\{r_1,r_2,\cdots,r_{l_r}\right\}$ and the specified music style s, the target generates a multi-track music with style s $A=\left\{\left(M_1',R_1'\right),\cdots,\left(M_t',R_t'\right)\right\}$, where i represents the i th track. The study defines the problem of generating traditional Korean music that incorporates the Middle Kingdom style in the following formalization:

Given the rhythm $R=\left\{r_1,r_2,\cdots,r_l\right\}$ and the melody $M=\left\{m_1,m_2,\cdots,m_{l_m}\right\}$, using the MSMICA model to control the style of music, thus generating multi-track music with the style of the Central Plains, while ensuring the harmony between the multi-track music.

Input: the melody M of the music, the corresponding rhythm R, and its style s.

Output: a sequence of multiple tracks $A=\left\{\left(M_1',R_1'\right),\cdots,\left(M_t',R_t'\right)\right\}$, where i stands for the i th track.

Some important mathematical notation is shown below:

M is the sequence of melodies in the music.

R is the sequence of rhythms in the music.

C is the chord progression in the music.

A is the sequence of arrangements in the music.

p_i is the i th phrase in the music.

$m_{i,j}$ is the j th note in the i th phrase in the music.

$r_{i,j}$ is the duration of the j th note in the i th phrase of the music.

c_i is the i th chord in the chord progression.

l_m, l_r, l_c is the length of the melody, rhythm and chord progression.

$\bar{h}_{i,j}^m, \bar{h}_{i,j}^r, \bar{h}_{i,j}^c$ are the j th hidden state of the i th phrase of the melodic, rhythmic and chordal progression.

$h_{t,k}^i$ is the hidden state of the i th task at the k th step in the t th phrase.

3.2 Multi-style melody and arrangement generation modeling

The Multi-Style Melody and Arrangement Generation (MSMICA) model is shown in Figure 1, and the model contains the following main components:

Input: the rhythmic sequence $R=\left\{r_1,r_2,\cdots,r_l\right\}$, whose corresponding melody $M=\left\{m_1,m_2,\cdots,m_{l_m}\right\}$

Data Processing Section: The collected music data are subjected to a number of pre-processing operations, including melody extraction.

Generator part: given the rhythm R and melody M, the MICA model is used to generate multi-track harmonized music.

Discriminator section: Two discriminators are designed, i.e., multi-style discriminator and harmony discriminator.

Output and Generated Music Section: A sequence of Korean traditional music incorporating elements from the Central Plains is generated and converted into a musical score, which is performed by a series of traditional instruments.

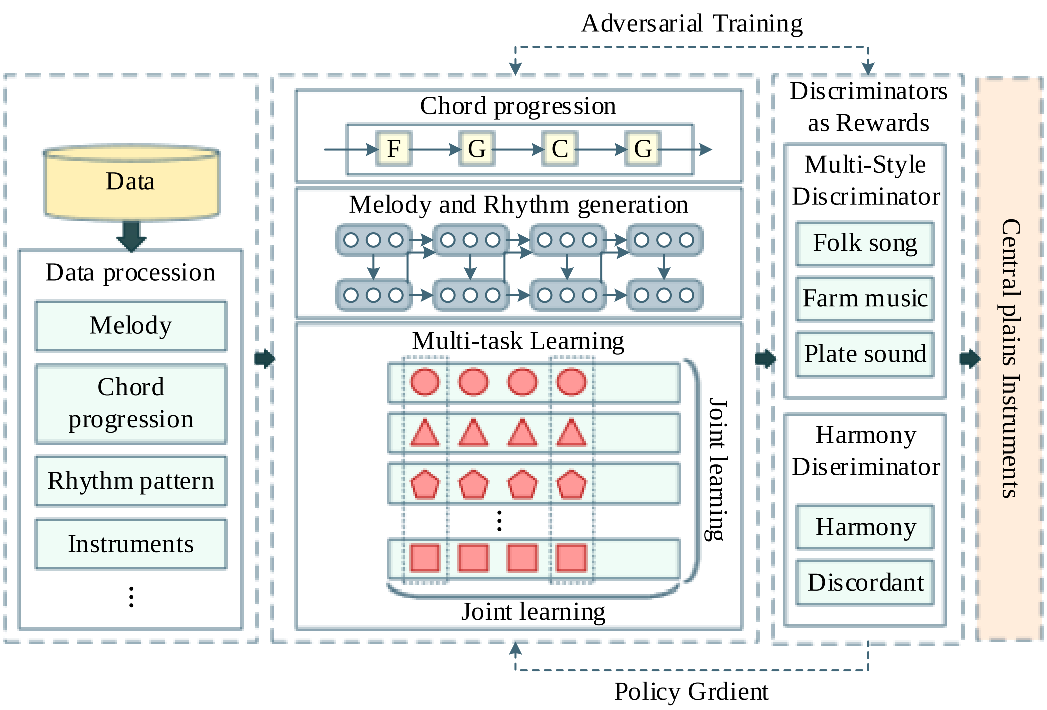

Figure 1: Multi-style melody and choreography model

3.2.1 Multiple Sequence Generative Adversarial Networks

The research utilizes an unsupervised Generative Adversarial Network (GAN) to control parallel data of mesogenic music styles, and the GAN usually utilizes a discriminator model to guide the training of the generative model, so that the generative model is constantly close to the real data. Containing the music style S music generator G and discriminator model D as follows:

$$\min_G \max_V V(D,G) = E_{s\sim p_{data}(s)}\left[\log D(s)\right] + E_{x\sim p_x(x)}\left[\log\left(1-D\left(G(x)\right)\right)\right] \tag{4}$$

$p_x(x)$ is a piece of music x generated by a generator G, which is compared to real music $p_{data}(s)$ based on the style S, which can be defined as a two-player zero-sum game with value function $V(D,G)$. Since the sequence is discrete, it is not possible to pass the gradient directly from the discriminative model to the generative model. To solve the above problem, Multi-Seq GAN is proposed to process multiple sequences simultaneously with a new sampling and rewarding method, and then use multiple discriminators to control the music style and harmony.

3.2.2 Generator

Considering the relationship between multi-tracks, the model chooses MICA as the generator to generate traditional Korean music in the style of Middle Chinese music. The modeling process can be divided into the following three parts:

(1) Sampling

Considering that the discriminator provides rewards only after completing the sequence, the generator utilizes the Monte Carlo search sampling strategy G_β to sample the unknown final $T-t$ word block:

$$\left\{Y_{1:T}^{1}, \cdots, Y_{1:T}^{N}\right\} = MC^{G_\beta}\left(Y_{1:t}; N\right) \tag{5}$$

Based on this method, multiple sequence sampling is defined as follows:

$$Y^{s} = \begin{bmatrix} Y_{1:1}^{1} & Y_{1:2}^{1} & \cdots & Y_{1:T}^{1} \\ Y_{1:1}^{2} & Y_{1:2}^{2} & \cdots & Y_{1:T}^{2} \\ \vdots & \vdots & & \vdots \\ Y_{1:1}^{i} & Y_{1:2}^{i} & \cdots & Y_{1:T}^{i} \end{bmatrix}, \; S_t = Y_{1:t}^{1} \oplus Y_{1:t}^{2} \oplus \cdots \oplus Y_{1:t}^{i} \tag{6}$$

where $Y_{1:t}^{i}$ is the sampling result of the i th sequence in step t, which represents the sampling of an unknown $T-t$ word block based on the existing t word block sequence. Considering the relationship between tracks, multiple sequences are concatenated into the generated data S_t, where t denotes t-step sampling.

(2) Reward

Since the generator cannot be rewarded directly, based on reinforcement learning, the model uses the feedback from the discriminator $D_\phi(S_t)$ as the reward. In this case, the reward of the strategy gradient uses the sum of all valid action sequences to anticipate future rewards:

$$Q_{D_\phi}^{G_\beta}\left(s = S_{t-1}, a = a_t\right) = \begin{cases} \dfrac{1}{N}\sum_{n=1}^{N} D_\phi(S_T^{n}), & for\ t < T \\ S_T^{n} \in MC^{G_\beta}(S_t; N) & \\ D_\phi(S_t) & for\ t = T \end{cases} \tag{7}$$

$$a_t = y_t^{1} \oplus y_t^{2} \oplus \cdots \oplus y_t^{i}$$

where $Q_{D_\phi}^{G_\beta}$ is the reward of the discriminator D_ϕ at step t. S_{t-1} includes the existing multiple sequences at step $t-1$, and a_t is the t th step operation which has multiple sequence operations. The reward matrix is defined as follows:

$$G(Y^s)=\begin{bmatrix} G(Y_{1:1}^1) & G(Y_{1:2}^1) & \cdots & G(Y_{1:T}^1) \\ G(Y_{1:1}^2) & G(Y_{1:2}^2) & \cdots & G(Y_{1:T}^2) \\ \vdots & \vdots & & \vdots \\ G(Y_{1:1}^i) & G(Y_{1:2}^i) & \cdots & G(Y_{1:T}^i) \end{bmatrix}$$
$$=\begin{bmatrix} G_{D_\phi}^{G_\beta}(S_0,a_1) & G_{D_\phi}^{G_\beta}(S_1,a_2) & \cdots & G_{D_\phi}^{G_\beta}(S_{T-1},a_T) \\ G_{D_\phi}^{G_\beta}(S_0,a_1) & G_{D_\phi}^{G_\beta}(S_1,a_2) & \cdots & G_{D_\phi}^{G_\beta}(S_{T-1},a_T) \\ \vdots & \vdots & & \vdots \\ G_{D_\phi}^{G_\beta}(S_0,a_1) & G_{D_\phi}^{G_\beta}(S_1,a_2) & \cdots & G_{D_\phi}^{G_\beta}(S_{T-1},a_T) \end{bmatrix} \tag{8}$$

(3) Loss function

The generator is defined as the whole environment, each word block is defined as an action, and the output of the discriminator is used as reward feedback. To boost the reward value of the whole environment, the model maximizes the sampling reward as follows:

$$J(\theta)=\sum_{i\in I}\sum_{y_{1:T}^i\in Y} P_\theta\left(y_{1:T}^i\right)R\left(y_{1:T}^i\right)=\sum_{i\in I}\left(E_{y_{1:T}^i\sim p_\theta}\sum_{t=1}^{T} r\left(y_{1:t}^i\right)\right) \tag{9}$$

where i is defined as the i th sequence, $r\left(y_{1:t}^i\right)$ represents the reward received at moment t, and $R\left(y_{1:T}^i\right)$ represents the cumulative reward for the previous T time period.

3.2.3 Discriminators

Two important requirements need to be met in order to generate Korean traditional music with a centralized musical style:

1) A specific musical style should be controlled.

2) The harmony of the generated multi-style music should be maintained.

(1) Style discriminator

This discriminator is used to control the style of the generated music, and this module is utilized to identify the music of the Middle Kingdom style. The hidden state c of the multi-track music sequence y is obtained using GRU and then the classification result is obtained using multilayer perceptual network and SoftMax function. The process is as follows:

$$\begin{gathered} c=GRU\left(y\right) \\ f=\tanh\left(W_c\odot c+b_c\right) \\ C_m=soft\max\left(W_m\cdot f+b_m\right) \end{gathered} \tag{10}$$

where W_c,b_c,W_m,b_m are the parameters to be trained, $\odot$ is the dot-product operation, and C_m is defined as the probability of the individual style discriminations in the multi-style discriminator.

(2) Harmony discriminator

Using the style discriminator, the style of the generated music can be controlled by fine-tuning the strategy gradient, but the harmony of the resulting music may not be maintained. The harmony discriminator guides the generated music towards harmonic music, and the generated music is categorized into three types: harmonic, generative, and dissonant. Harmonious music is considered as positive example samples and other music is defined as negative example samples, GRU and fully connected layers are utilized to obtain the hidden state of the music sequence Y:

$$C_h = softmax\left(W_h \odot GRU(y) + b_h\right) \tag{11}$$

3.2.4 Model training

The outputs of the two discriminators are used as feedback to train the MSMICA model, while the feedback function is defined as the linearly weighted gradient value of the classification effect of the two discriminators, which is balanced by the λ parameter as follows:

$$R(y|\cdot) = \lambda C_s\left(c = style|y\right) + (1-\lambda) C_h\left(c = harmony|y\right) \tag{12}$$

Due to the adversarial training approach, the problem is defined as the zero-sum function $V(G,D)$ of the maximum and minimum values of the generator G and the discriminator D:

$$\min_G \max F\left(V(D_1, G), \cdots, V((D_n, G))\right) \tag{13}$$

where $n = 2$ and F is defined as the linear weighting of the discriminator. The generator G needs to be pre-trained to pass the maximum likelihood function before training with discriminant gradient feedback.

3.3 Composition process

The MSMICA model is continuously trained to learn and eventually a converged good note prediction model is obtained. Using this model it is possible to generate a fixed-length sequence of notes, i.e., a compositional process.

The note prediction model also requires the input of a fixed length of n musical note sequences, the first input note sequences are randomly selected from the test set, and then after the note prediction model outputs the predicted note sequences of the next fixed length of n musical note sequences, and then the output of the predicted musical note sequences as the input to carry out the next prediction, and so on iterating until the generation of pre-set generated musical note length, forming an automatic composition, the above is a good automatic composition. The above is the whole process of generating a piece of music by the automatic composition model. In the automatic composition model, the input data set of the note prediction model to generate a piece of music can be expressed as follows:

$$X = \begin{bmatrix} note_1 & note_2 & \cdots & note_n \\ note_{pre_1} & note_{pre_2} & \cdots & note_{pre_n} \\ \vdots & \vdots & & \vdots \\ note_{pre_{i*n+1}} & note_{pre_{i*n+2}} & \cdots & note_{pre_{(i+1)*n}} \end{bmatrix} \tag{14}$$

The note sequences of generated musical notes output by the note prediction model can be expressed as the following equation:

$$Y=\begin{bmatrix} note_{pre_1} & note_{pre_2} & \cdots & note_{pre_n} \\ note_{pre_{n+1}} & note_{pre_{n+2}} & \cdots & note_{pre_{2*n}} \\ \vdots & \vdots & & \vdots \\ note_{pre_{(i+1)*n+1}} & note_{pre_{(i+1)*n+2}} & \cdots & note_{pre_{(i+2)*n}} \end{bmatrix} \tag{15}$$

The first line of data $X_1=\left[note_1, note_2, \cdots, note_n\right]$ in X denotes the randomly selected sequences of musical notes from the test set for each generation of a musical composition. The first line in the corresponding Y is the sequence of musical notes it generates $Y_1=\left[note_{pre_1}, note_{pre_2}, \cdots, note_{pre_n}\right]$. Then Y_1 is assigned to X_2 as the input for the next prediction, and sequential iterations are carried out for the generation of musical note sequences, and finally the musical note sequences generated in Y are spliced together by rows to be the complete note sequences of the generated musical composition. Figure 2 illustrates the composition process of the MSMICA model.

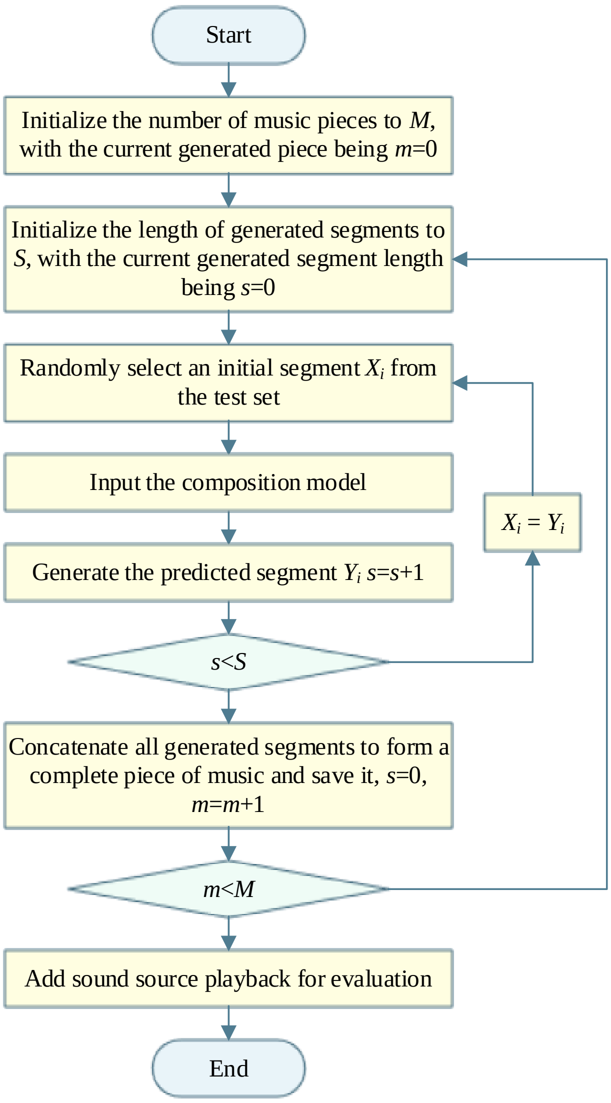

Figure 2: MSMICA model composition process

4 Re-creation of Korean traditional music by integrating elements of music from the Central Plains

In this section, the performance of the above model and the quality of the generated music are experimentally verified. The experimental environment is as follows:

CPU: Intel(R) Xeon(R) CPU E5-2678 v3 @2.50GHz
Memory: 256GB
Graphics processor: TITAN RTX
Operating system: Ubuntu 18.04.2 LTS

4.1 MSMICA model training process

The study uses the training error and the accuracy of music classification in the Central Plains style to validate the feasibility of the MSMICA model. The effectiveness of the trained model is verified by introducing a loss function and a correctness prediction function in the training process. Figure 3 shows the training iteration process of the MSMICA model, and the Loss value gradually leveled off around 50 rounds of iteration, and finally stabilized at around 0.44. The accuracy of the model in classifying music in the Central Plains style shows drastic fluctuations between 20 and 60 rounds of training, and stabilizes after 60 rounds, with the accuracy converging to 0.95 or less. Within 65 rounds of iteration, the loss function curve and the accuracy curve flatten out, which is in line with the iterative law of the deep learning curve, which indicates that the model obtained from the research training is effective.

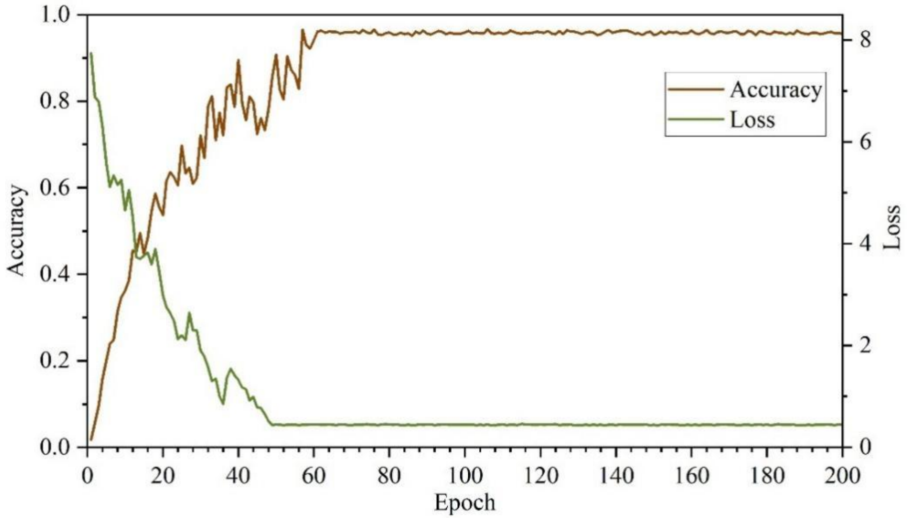

Figure 3: The training iteration of the MSMICA model

4.2 Multi-perspective assessment of the quality of generated music

In order to test whether the MSMICA model is effective in generating traditional Korean music in the Middle Kingdom style, the quality of the generated music needs to be evaluated. In order to test the music generation effect of the algorithm, this section evaluates the quality of the model-generated traditional Korean music in the Central Plains style based on the signal characteristics, Meier spectrum, spectral rolloff, and chromatographic frequency of the generated music.

(1) Signal characteristics

Figure 4 shows the waveforms of the model-generated music samples, and it can be seen that the signal fluctuation carries a certain regularity, and the amplitude of the fluctuation can be kept within a certain range (±0.8), which indicates that it is possible to learn the internal laws of the music data using the MSMICA model.

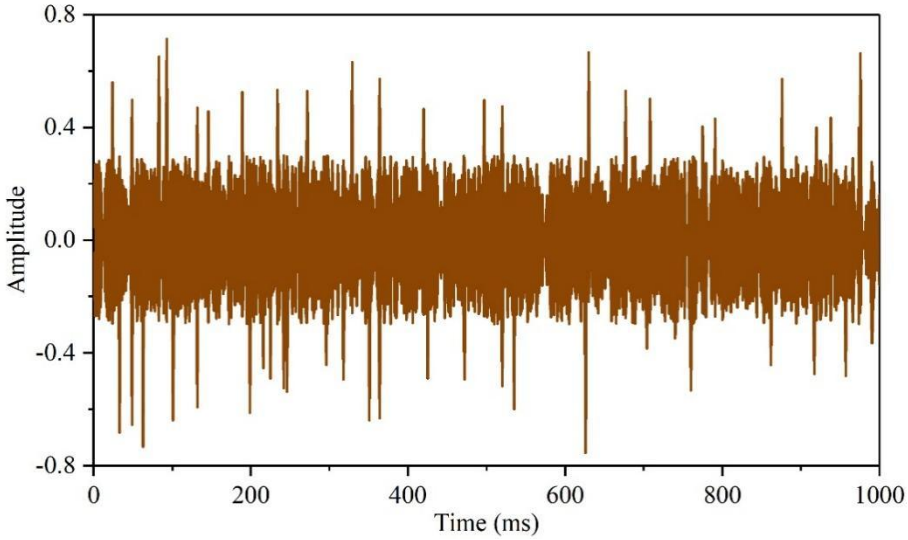

Figure 4: The model generates a waveform of the music sample

(2) Mel Spectrum

Figure 5 shows the Mel spectrum of the generated audio.The MSMICA model generates a relatively tight tempo of the traditional Korean music in the Central Plains style, and the frequency basically maintains the fluctuation within a certain interval.

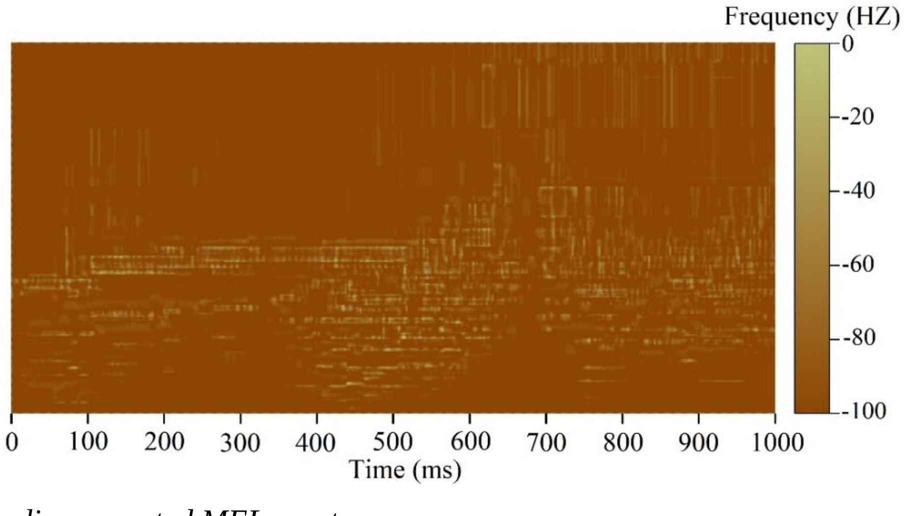

Figure 5: Audio generated MEL spectrum

(3) Spectral roll-off

The spectral roll-off of the generated audio is shown in Figure 6. The spectral decay emphasizes the degree of rhythmic change, and it can be seen that the music's spectral decay carries a certain rhythm that is more stable.

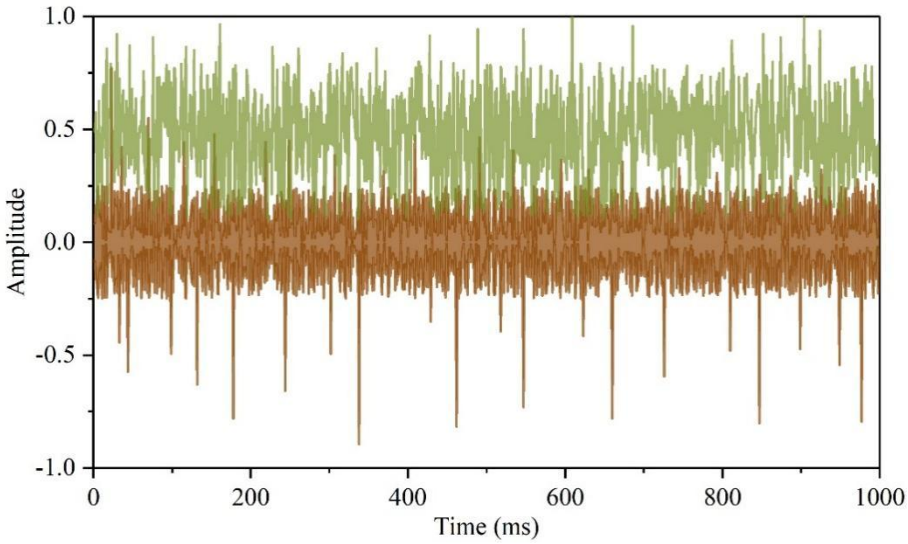

Figure 6: The spectrum of the generated audio is rolled down

(4) Chromatographic Frequency

Figure 7 illustrates the chromatographic frequencies of the generated audio. The chromatographic frequencies are the 12 different semitones (or chromatic degrees) that represent the octaves of the music. The traditional music of the Central Plains style of Korea generated by the MSMICA model has a better representation of the variation of pitches and the presentation of melody and rhythm in terms of the chromatographic frequencies.

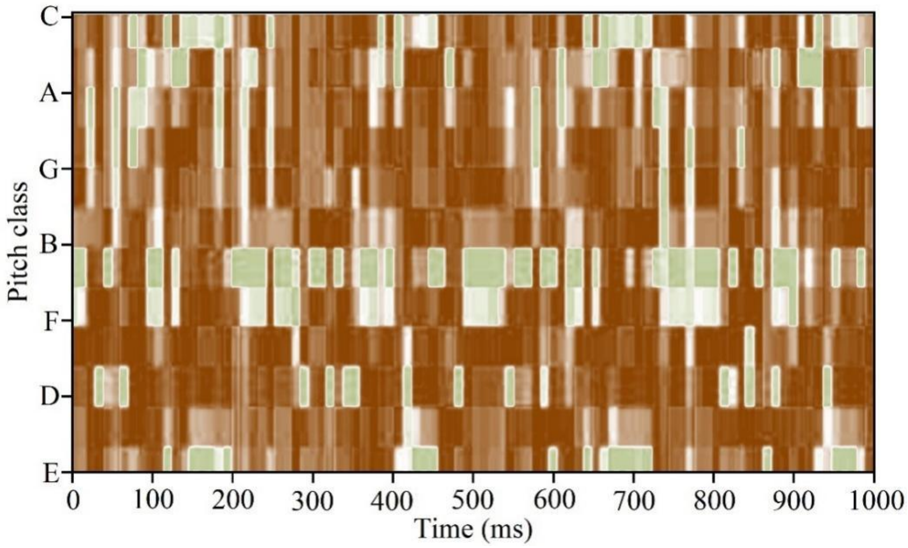

Figure 7: Generate the frequency of audio chromatography

By comparing the audio signal features such as Mel spectrogram, spectral attenuation, chromaticity frequency, etc. of the traditional Korean music in the Central Plains style generated by the MSMICA model with the original music, it can be found that there is basically no obvious difference between the signal features of the generated music and the original music, which can be seen in the strong applicability of the MSMICA model.

4.3 Comparison of music generation with different models

In order to verify the feasibility of the MSMICA model in generating the melody of Korean traditional music in the Central Plains style, it is necessary to compare and evaluate the generated results, and here Transformer and GAN models are used to do a comparative analysis of the melody. By using the same music text data to train the Transformer and GAN models, different models are obtained to generate Korean traditional music segments incorporating elements of music from the Middle Kingdom, of which Fig. 8~Fig. 10 show the music segments generated by the MSMICA, Transformer and GAN models, respectively.

As can be seen from the green arrows in the figures, which indicate the trend of the musical melodic progression, the melodic fragments of Korean traditional music in the Chungwon style generated using the MSMICA model and the Transformer model show a melodic progression with high and low levels. The melodies roughly fluctuate up and down around the main tone of the fragment, and then return to the main tone at the end, which is in line with the melodic progression laws of Korean traditional music and music of the Central Plains. The music melody fragments generated using the GAN model, on the other hand, show an overall downward trend, with the melody going high and low, which is not quite in line with the compositional laws of Korean traditional music and Middle Kingdom music, and sounds like a low and boring and unstable feeling. It shows that the MSMICA model and Transformer model are more capable of producing long sequences of musical melodies compared to the GAN model in model training.

In the above three different models for generating musical melodies, the comparison continues with the variation of notes as shown in the black boxed part of the melody in the figure. In the musical melody segments generated by the MSMICA model, the relative degree of note variation is small, and the melody proceeds slowly, with regular ups and downs around the dominant note. In the melodic fragments generated by the Transformer model and the GAN model, on the other hand, the note changes are too aggressive, sometimes spanning close to an octave. Such a melody sounds abrupt and does not conform to the compositional rules of Korean traditional music and Middle Kingdom music. After comparison, the traditional Korean music of the Central Plains style generated by the MSMICA model is closer to real music compared to other models, and the model has high accuracy and the resulting music has the characteristics of real music.

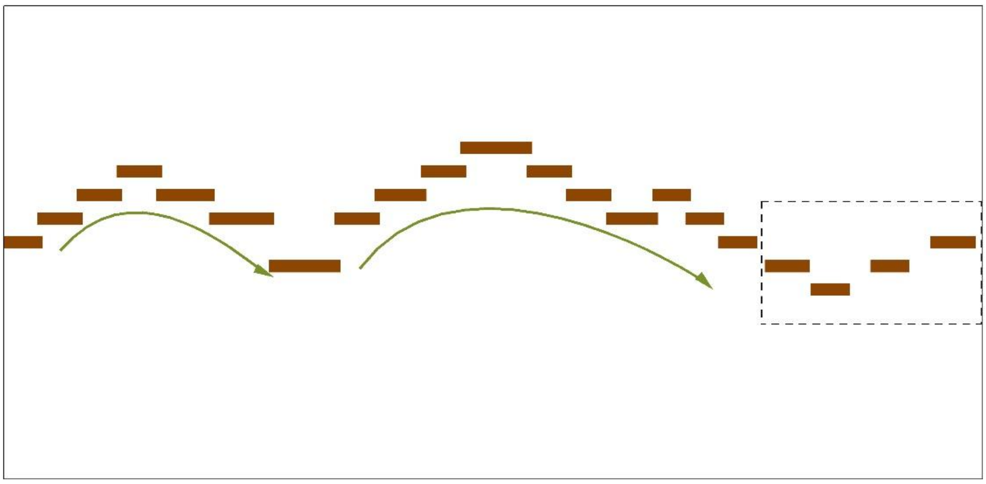

Figure 8: MSMICA model generated music segment

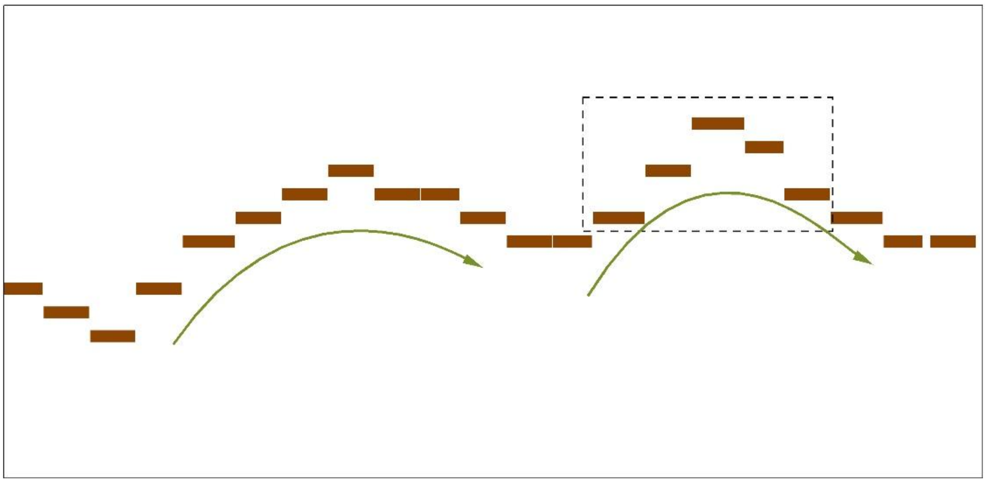

Figure 9: Transformer model generated music segment

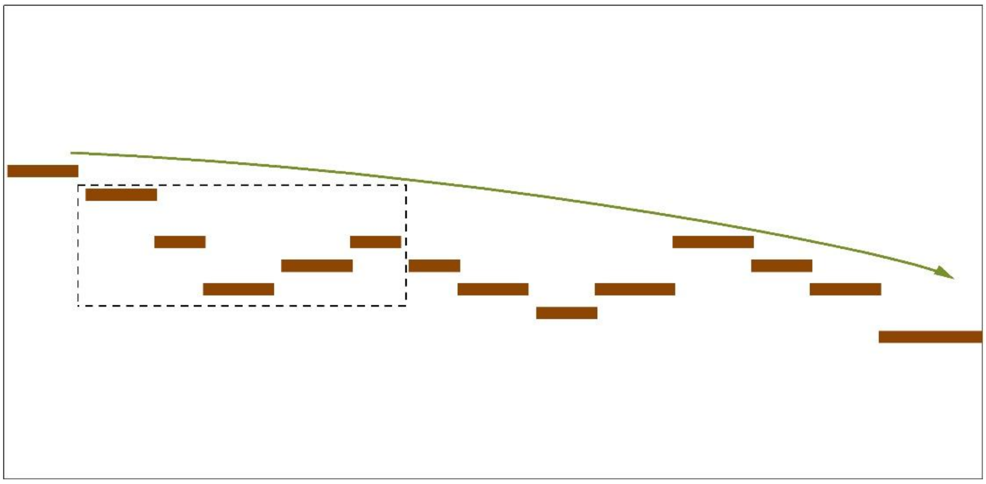

Figure 10: GAN model generated music segment

4.4 Subjective and comprehensive evaluation of the main theme

The generation of traditional Korean music in the Central Plains style was carried out according to the specific method described above, and in order to ensure the fairness of the evaluation, five relevant melodies were generated by each of the three models MSMICA, Transformer and GAN. Each model uses the dataset collected above and selects the same parameters to control the key, tempo, beat number and other information of the generated melodies to evaluate the five randomly generated melodies. To ensure the accuracy of the evaluation, the study conducted a subjective and comprehensive assessment of the generated pieces in terms of the melody's neutrality, structure, pleasantness, association, and resonance. Since aesthetic differences exist objectively, 20 students each from music majors and non-music majors were invited to give ratings to the generated melodies in each of these five areas. The value of the comprehensive evaluation = the average value of the music major's evaluation * 0.7 + the average value of the amateur evaluation * 0.3.

The score of subjective evaluation was set from 1-10, and the results of the subjective comprehensive evaluation are shown in Table 1, where the values are averages. In terms of professional music expertise evaluation, the MSMICA model improved 12.34% to 35.75% over the Transformer and GAN models in the five subjective evaluation indexes, while in terms of amateur evaluation, the MSMICA model improved 10.36% to 35.36% over the comparison models. Overall, the MSMICA model showed an improvement of 11.63% to 35.67% over the Transformer and GAN models in the overall subjective composite evaluation value, with the increase in the neutrality metrics ranging from 28.47% to 35.67%, which showed a more significant improvement.

The MSMICA model proposed in the study is advantageous in the generation of traditional Korean music of the Central Plains style, as shown in:

(1) Clear tuning: the melodic tuning generated by the MSMICA model was Gong, Levistic, and Symbolic, which conformed to the tonal characteristics of the Chungwon style of music, while the tuning of the melody generated by the comparison model was not obvious.

(2) Accurate intervals: The MSMICA model is more accurate than the Transformer and GAN models in grasping the intervals of the music of the Central Plains style, and the melodies generated are more characteristic of the Central Plains.

(3) Rhythmic regularity: The rhythmic regularity of the melodic phrases generated by the MSMICA model, and most of the phrases in each melody end with a short rhythm into a long rhythm. On the other hand, the rhythms of the melodies generated by the Transformer and GAN models are more scattered, and it is difficult to have a clear grasp of the length of the phrases.

Table 1: Subjective comprehensive evaluation results

		Central	Structure	Melodious	Association	Resonance
Professional	MSMICA	8.62	8.77	9.11	8.43	8.56
	Transformer	6.71	6.76	7.42	7.18	7.57
	GAN	6.35	7.19	6.88	6.66	7.62
Amateur	MSMICA	8.92	9.01	9.5	8.72	8.84
	Transformer	6.94	7.02	7.72	7.44	7.81
	GAN	6.59	7.44	7.11	6.96	8.01
Synthesize	MSMICA	8.71	8.84	9.23	8.52	8.64
	Transformer	6.78	6.84	7.51	7.26	7.64
	GAN	6.42	7.27	6.95	6.75	7.74

The global structural nature of the main melody of the model-generated Central Plains style Korean traditional music was analyzed again in conjunction with the melodic line. Figure 11 shows the melodic line trend of the MSMICA model-generated music segments. The correlation between the phrases within each segment is high, the second phrase is a repetition or modal progression of the first phrase, the third phrase exhibits a transitive function and is the climax of the whole segment, and the fourth phrase serves as a summation, which is a typical starting and ending structure.

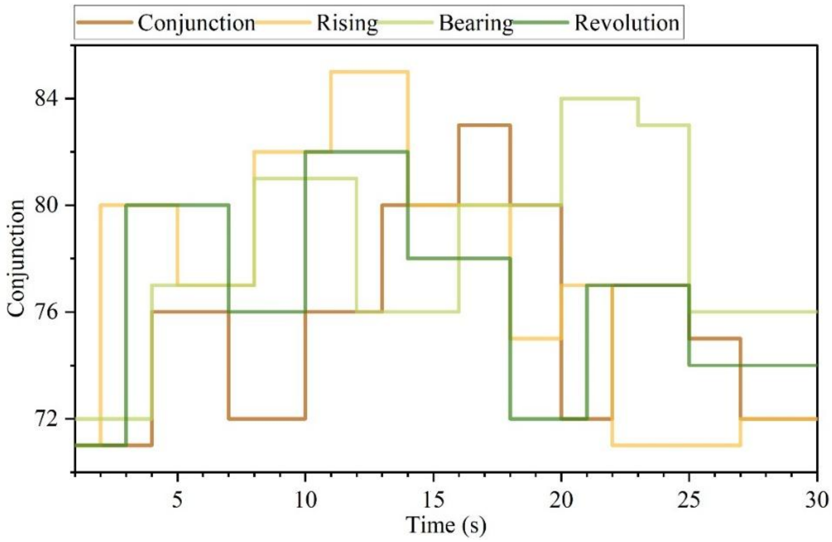

Figure 11: The MSMICA model generates the melody line of the music segment

5 Conclusion

This research extracts the melodic features and constructs the note feature vectors in the music data of the Middle Kingdom, and then utilizes the Korean traditional music generation model based on the multi-sequence generative adversarial network to deeply learn the feature elements of the Middle Kingdom in order to realize the absorption and recreation of the Middle Kingdom elements in the Korean traditional music.

The model obtained from the research training can effectively realize the fusion of the Middle Kingdom music elements in the Korean traditional business, with fast training speed and good convergence value.

The signal characteristic fluctuation range of the music generated by the MSMICA model is between ±0.8, and the Mel frequency is also basically maintained in a certain range, showing strong application value.

The model-generated fragments of traditional Korean music in the Central Plains style showed a trend of high and low ups and downs in the melodic direction, and the note changes were more gentle, which is in line with the characteristics of music in the Central Plains.

In terms of subjective comprehensive evaluation, the comprehensive scoring values of the music fragments generated by the model in terms of the indicators of Middle Kingdom, structure, etc., were improved by 11.63%~35.67% compared with the comparison model, and the comprehensive scores in terms of the indicators of Middle Kingdom were improved by 28.47%~35.67% compared with the comparison model. The model has obvious advantages in the absorption of Central Plains musical elements and can accomplish the task of multi-style music fusion.

About the Author

Zhentang Cheng was born in Xuzhou, Jiangsu, P.R. China, in 1990. He obtained a bachelor's degree from Jiangsu Normal University, a master's degree from Qufu Normal University, and a doctorate from Yeungnam University in South Korea. He is currently a postdoctoral researcher at Anhui Normal University. His main research focus is the history of modern and contemporary Chinese music.

References

[1] Kong, W. (2025). A comparative study of court music education between Ming Dynasty and Korea Dynasty. Mediterranean Archaeology and Archaeometry, 25(2).

[2] Guo, L. (2022). Court music of China, Japan and Korea in the 16th-17th centuries: a comparative analysis. Musica Hodie, 22.

[3] Noh, A. S. (2019). Cultural policies for national music in South and North Korea (1960s–70s): a comparative study. International Journal of Cultural Policy, 25(1), 20-32.

[4] Kim, J. H. (2019). A Study on the Musical Characteristics of Fishing songs in North Korea. (The) Research of the performance art and culture, (39), 279-323.

[5] Peter, M. O. O. D. Y. (2021). Reconstructing the National Heritage: Socialist Folk Music in North Korea and East Germany, 1945–1963. Korea Journal, 61(1), 186-218.

[6] Moody, P. (2024). Life Songs and Lifelong Expertise: The Hidden Sustainable Properties of Music in North Korean Secondary Cities. In Pursuing Sustainable Urban Development in North Korea (pp. 48-58). Routledge.

[7] Cloonan, M. (2021). Songs for 'Great Leaders': Ideology and creativity in North Korean music and dance. By Keith Howard. Oxford: Oxford University Press, 2020. 349 pp. ISBN 978-0-19-007751-8. Popular Music, 40(1), 186-189.

[8] Li, X. (2017, June). A Study on the Characteristics and Transmission Ways of China Central Plain Opera Culture. In 2017 2nd International Conference on Education, Sports, Arts and Management Engineering (ICESAME 2017) (pp. 1-4). Atlantis Press.

[9] Zuo, Y. (2020). Keeping Your Ear to the Cosmos: Coherence as the Standard of Good Music in the Northern Song. In Powerful Arguments (pp. 277-309). Brill.

[10] Shang, C. (2024). The Archaeological Value of Tang Dynasty Music Singing and its Influence on the Innovation of Ancient Poetry in Jiangnan. Mediterranean Archaeology and Archaeometry, 24(2), 250-262.