International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-15

https://doi.org/10.70917/ijcisim-2025-0301
Article

Research on the Innovative Inheritance Application of Artificial Intelligence Technology in Traditional Music Education and Strategies to Enhance the Aesthetic Experience

Ning Li *

Educational and Scientific School, Zhengzhou Normal University, Zhengzhou, Henan, 450000, China; lining_1123_1123@163.com

Abstract: With the deep integration of digital technology and traditional art, artificial intelligence (AI) provides new momentum for the innovative development of traditional music education. This paper focuses on the cultural inheritance and aesthetic nurturing demands of traditional music education, systematically analyzes the inheritance and innovation application strategies of AI technology in traditional music education, constructs a music automatic generation model based on Leak-GAN, and conducts experimental investigations on students' aesthetic sports in music education. The results of the study show that the music melodies generated by the Leak-GAN model in this paper have obtained general listeners' ratings of 4.1, 4.1, and 3.9 in the aspects of coherence, pleasantness, and innovativeness, respectively, and professionals' ratings of 3.9, 4.1, and 3.5, which are better than those of other comparative models. The expert group students scored significantly higher than the novice group in music aesthetics, emotional experience, and music-induced arousal, indicating that music training can regulate the aesthetic judgment of music, enhance the pleasurable experience, and is an important part of enhancing the effect of music aesthetics. This paper is of great practical significance for the inheritance and innovation of AI technology in music education and the development of strategies to enhance students' aesthetic experience.

Keywords: traditional music education; cultural heritage and innovation; Leak-GAN; music automatic generation

1. Introduction

Traditional music is an artistic treasure in the treasury of human culture, which not only carries the memory of history, but also embodies the characteristics and emotions of the Chinese nation. In China, traditional music includes court music, folk music, religious music and literati music, etc. The content can be subdivided into folk songs, instrumental music, rap music, opera music, dance music, etc. [1-3]. These distinctive forms of music together constitute a rich and colorful traditional music culture, which is an important carrier of Chinese cultural heritage and national identity [4]. However, limited by geographical differences and the imbalance of educational resources, traditional music education in China has always been faced with the real problems of lack of teaching resources, insufficient teachers, and relatively single teaching mode, which makes it difficult to meet the needs of modern education [5-6]. And the emergence of artificial intelligence (AI) technology has brought new opportunities and challenges to traditional music education.

With the development of AI, the emergence of virtual classrooms and online learning platforms provides new ways to expand teaching resources. Literature [7] points out that the advantage of the music education network virtual classroom platform lies in its openness, scalability, flexibility, and near-permanent distance teaching separation, which enable the dissemination of traditional music teaching resources to the outside world.AI can break through the limitations of time and space, enabling students to have access to excellent traditional music teaching resources anytime and anywhere [8]. For example, with the help of AI-assisted virtual classrooms, students from remote areas can receive the same quality of traditional music teaching as students in big cities. Meanwhile, by combining VR (virtual reality) and AI technologies, an immersive teaching experience is realized. For example, literature [9] created an interactive music course teaching model by using VR and AI technologies, and verified the feasibility of combining VR and AI technologies with music teaching through experimental analysis. As a result, students can not only learn the repertoire through video, but also interact with the virtual teacher to obtain personalized listening, singing, and playing guidance, providing a broader platform for the development of traditional music education.

The application of AI in traditional music education provides a customized learning experience for each student, and this personalized learning is mainly reflected in the fact that AI can analyze students' learning performance in real time, identify their technical deficiencies and give targeted feedback [10]. For example, literature [11] developed and designed a music education information system based on multimodal fusion and reinforcement learning, which is capable of dynamically recommending personalized learning paths and resources through real-time collection and analysis of multimodal data such as students' music fundamentals, learning preferences, and emotional states. Literature [12] constructed an intelligent music education system (MES) based on machine learning algorithms and AI technology, which is used to automatically generate music accompaniment through intelligent generative adversarial network (IGAN), and optimize the recommendation process of personalized learning paths with the help of Adam's optimization algorithm (RAO). Through personalized learning, students are able to experience more diverse learning styles, increase their interest and motivation, and improve the efficiency of traditional music learning.

Taking intelligent learning platforms as an example, certain ethno-instrumental music teaching platforms have realized real-time analysis and feedback on students' performance. For example, literature [13] points out that functional digital musical instruments (DMI) bring different haptic feedback to players through human-computer interaction technology, and provide comprehensive data to quantify the feedback effect of haptic interaction. This personalized feedback mechanism not only enables students to improve their playing skills in a shorter period of time, but also helps teachers to achieve a higher quality of tutoring within the limited teaching time. Literature [14] constructed a dynamic model of Sound Performance-Subject Driven Environment (SSE) to optimize traditional Chinese musical instruments, and they believed that the digital transformation and sound optimization of traditional Chinese musical instruments is the main force to promote their internationalization, while professional knowledge and aesthetic pursuit play a crucial role in it. Literature [15] takes the teaching of traditional Chinese musical instrument erhu as an example, pointing out that flipped teaching (CPA) combined with AI technology can analyze the problems of playing angle and strength, and give specific operation suggestions such as adjusting the bow speed and the bowstring contact angle, and at the same time, it can also reduce the students' nervousness when playing. In summary, AI technology empowers traditional music education, especially in the popularization of educational resources, personalized learning and teaching feedback and other aspects of the huge potential, but also faces the imbalance between technology and art [16], the reliance on the role of the teacher [17], the lack of depth of student learning [18] and other aspects of the challenge. Therefore, further in-depth research on it is urgently needed.

This paper explores the strategy of using AI technology in music education to carry out innovative inheritance application and enhance the aesthetic experience. Based on the knowledge of music theory, the construction of a music generation model is realized through Leak-GAN, so as to realize the intelligent generation of music. In order to verify the practicality of the proposed model, the model was subjected to listening and comparison experiments. On this basis, experiments were designed to investigate the effects of music emotion type, music preference and music training experience on students' aesthetic experience, providing suggestions and references for improving students' aesthetic experience.

2. Innovative inheritance application of artificial intelligence technology in traditional music education

This chapter provides a brief introduction to strategies for integrating artificial intelligence technology for heritage application innovation in teaching rhythmic and melodic composition in traditional music education.

2.1. Encourage students to create rhythms

In music teaching, after teachers carry out some theoretical knowledge learning of rhythm for students, they can take advantage of artificial intelligence technology to encourage students to boldly create rhythm and melody. The creation of rhythm and melody learning is mainly divided into two forms, one is to use the music material as the basis for its adaptation or innovation, and the other is improvisation.

(1) Rhythmic Arrangement Based on Musical Materials

The learning of melody and rhythm creation is not a quick fix. Teachers should help students create rhythms and melodies in a step-by-step manner from easy to difficult. At the beginning of the process, teachers can use artificial intelligence technology to collect a comprehensive collection of musical materials, enrich the students' expansion of musical materials, and let students use their own imagination and knowledge of music to imitate and innovate. For example, when teachers teach students the rhythm and beat, they can let students make their own adaptations by moving the bar line and adding the rhythm, and make full use of artificial intelligence technology to play the adaptations for students to enjoy in the classroom.

(2) Encourage students to improvise in the form of games based on multimedia technology

No matter what age group students like to learn by playing games, teachers can integrate the creation of rhythm and melody with games, and use artificial intelligence technology as a carrier to effectively stimulate students' inspiration for creation. For example, when the teacher carries out the rhythm solitaire game, the students will be divided into five groups, and each group will create the rhythm accordingly, and the results will be played on the spot through the artificial intelligence technology, so that the students can make an evaluation of the results of each creation, and the group that gets the excellent creation can get a small prize. Through this small game not only cultivate the students' ability to create, but also effectively improve the students' memory ability, as well as sensitivity to music, and cultivate the students' sense of rhythm.

2.2. Refinement of improvisation and choreography teaching content and methods

(1) Perception of rhythm

When taking music lessons, students should feel the charm of music through the sense of rhythm and rhythmic movement and learn to express music. For example, in the song “Papaya Cha Cha Cha”, teachers can let students feel the rhythmic movement of music through music and innovate the mode of teaching. For example, according to the song sheet, let the students follow the music to beat freely, or when the lyrics appear cha-cha-cha-cha, let the students make clapping movements to express the rhythmic points of the music.

(2) Rhythm Filling

Teachers can also take a part of the sheet music inside a song and let students fill it out according to their own understanding. For example, according to the beat number, fill in the blank part of the rhythm in the score shown.

(3) Imitating Variations

In order to improve students' motivation, teachers can let students carry out imitation, they can let students listen to a small piece of music rhythm in the classroom, and then let students use their own imagination to make adaptations to the rhythm as a way to let students deeply understand what rhythm is. For example, teachers can carry out rhythmic demonstration and imitation activities, and let students do improvisation in the 1st and 3rd sections.

(4) Rhythm Quiz

Teachers can also take a rhythm point inside a song as a quiz question and let students do improvisation with other beats. For example, a small game in the classroom, the students are divided into several groups, each group elects a group leader, the leader of the group first randomly create a rhythm of 2 bars, and use the word “da” to sing out, the other members of the group will be one by one down to solitaire, see which group solitaire the largest number of people.

3. Automatic music generation model based on generative adversarial network

This chapter further refines the study of artificial intelligence techniques used in music education, namely, the construction of a music generation model, Leak-GAN, to realize the automatic generation of rhythms and melodies improvised by students in music teaching. On this basis, application experiments are conducted on the model.

3.1. Introduction to Basic Music Theory and MIDI

Before proceeding to the intelligent generation of music, this section will first introduce the unit of music: the note, and the two most important attributes of music: pitch and duration, which are also the musical attributes that are important to consider in the modeling process of music in this paper. This is followed by a description of how the Music Digital Interface (MIDI) represents music in a computer.

3.1.1. Units and attributes of music

A note is the smallest unit of music, and is used to record the progression of tones of different lengths. If we break up a melody, we will find that it is made up of several specific musical events, and each musical event can correspond to a specific note.

In pentatonic music, the differences in shape and position of the notes indicate the two most important attributes of a note - pitch and duration.

Pitch refers to how high or low the frequency of a note's sound is perceived when a human listens to it. Modern music prescribes scientific pitch notation, where pitch is indicated by a combination of note names and octave numbers. The note names are represented by a cycle of letters from C to B, commonly known as “do re mi fa sol la si”. The octave symbol is represented by a number from 0 to 8. The starting key of the piano is the note A in the 0th octave, i.e. A0, and every note cycle from C to B to the right of it will be raised by one octave, and the number of the corresponding octave symbol will be increased by one.

Duration, the duration of a note. In pentatonic music, the dots of the notes are not exactly the same shape, and this is because they represent different lengths of note duration. In the more common 4/4 rhythm, the quarter note is used as the base, which represents one beat, then the quarter note to its left represents twice its duration, which represents two beats, and so on. On the right side of the quarter note, whenever the note has one more note end, the duration of the note is reduced by one-half, e.g., the eighth note represents half of the quarter note, i.e., half a beat, and so on.

In addition, there are many other attributes of music, such as timbre and intensity. Tone color, which indicates the characteristics of the sound, is determined by the waveform of the music, sound pressure, frequency spectrum and other factors. Tone intensity, which indicates the objective physical strength of the sound, is determined by the amplitude of the vibration of the articulating body. The objectively identical tone intensity, when transmitted to the human ear, needs to be combined with factors such as pitch and timbre to determine the subjective loudness of the sound to the human ear.

3.1.2. Computer representation of tones

Music Digital Interface (MIDI) is the most widely used electronic communication protocol for transmitting music messages worldwide. Electronic keyboards and other electronic music devices can output digital music information created by musicians on the device, when a computer or other device receives a MIDI music file, it can be played in accordance with the various parameters in the file signals.MIDI music file consists of two parts: a unique header block and a number of track blocks.

The header block consists of three parts: the identifier string, 4 bytes long, the content of the “MThd” hexadecimal representation that is 4d 54 68 64, indicating that the file for the Chunk data structure of the MIDI file. The length of the header block data area is 4 bytes, usually the content is 00000006, which means that the length of the next header block data area is 6 bytes. The header block data area, usually 6 bytes in length, stores basic information about the MIDI file such as file format, number of tracks, time format and other information.

The header block is followed by a number of track blocks. The content of the track blocks consists of three parts: the identifier string, which is 4 bytes long, and is the hexadecimal representation of “MTrk”, i.e. 4d 54 72 6b, and the length of the track block data area, which is 4 bytes long, and shows the length of the track block that follows the specified one. Track data area stores the current track contains a sequence of MIDI events, these events are usually by the time difference and record the details of the performance of the MIDI message consists of two parts of the MIDI message on the timbre, melody, playing strength and other musical factors to set the parameters, so that when the computer reads the MIDI message, it will be aware of the event is the moment by the pitch of the pitch at what height.

3.2. Standard Generative Adversarial Networks

Generative Adversarial Networks (GANs) [19] design two models that fight against each other: a generator G that captures and tries to simulate the distribution of real data as much as possible, and a discriminator D that determines whether or not the samples are from the real data, so that the generator and discriminator fight against each other and rise up alternately, and ultimately generate the result that is false to be true.

The specific structure of the generative adversarial network is shown in Figure 1. The generator is a generative network model which receives random noise for generative tasks. The discriminator is a discriminative network model that receives both real data and generated data, it is responsible for making a judgment on whether the received data is real or not without knowing whether the data is real or not, and the output $D(x)$ represents the probability that the discriminator considers x to be real data. When the discriminator is already very capable, if the data generated by the generator still confuses the discriminator and prevents it from making a correct judgment, then we consider that the generator has learned the distribution of the real data. This state is considered to be the Nash equilibrium in game theory, i.e., both parties have maximized their benefits in the current state and will not change their strategies.

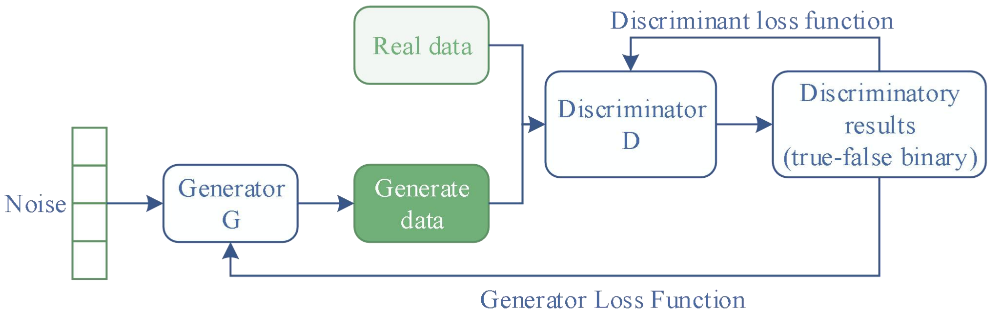

Figure 1. The structure of GAN.

The objective function for generating the adversarial network is:

$$\min_{G}\max_{D} V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log D(G(z))] \tag{1}$$

Here, the training discriminator D tries to judge the labels of the training samples correctly with maximum probability, i.e., maximizes $D(G(z))$ and $\log(1-D(G(z)))$. The training generator G minimizes $\log(1-D(G(z)))$ i.e. maximizes the loss of D. During the training process, each time one of the networks is fixed and the parameters of the other network are updated, alternating iterations, both sides try to maximize the other's error, and finally the generator G is able to generate distributions that are very close to the real data to achieve the generative effect.

3.3. Automatic Music Generation Based on Leak-GAN

In this paper, Leak-GAN model is chosen to generate music.Leak-GAN, mainly, is to leak the high-level discriminator feature information to the manager module, which uses the LSTM network to act as an intermediate medium to form a guiding behavioral target output, while the underlying generator inputs word vectors, which are passed through the worker module to get the behavioral vectors, and combined with the target output sampling to get the next word's output vector. The so-called Leak, which leaks the feature information of the discriminator into the generator and guides the generator generation, optimizes the traditional adversarial model.

For automatic music generation, the textual representation of melody approach studied in this paper allows music generation to be viewed as a one-dimensional sequence generation process. In each time step t, the current state of the model is denoted as $s_t = \left(x_1, \ldots, x_i, \ldots, x_t\right)$ and x_{t+1}, respectively, where x_i is a candidate token in the given lexicon V. Improvement of generator G_θ is guided by training θ parameters in generator G_θ and ϕ parameters in discriminator D_ϕ and using discriminator D_ϕ.

After allowing the high-level feature representation of the discriminator D_ϕ to leak to the generator G_θ, which should not receive such information, the Leak-GAN model solves the problem that the bootstrap signals of D_ϕ do not signal the complete care of generating the entire melodic sequence S_T. Specifically, as the length of the generated sentence T becomes longer, the bootstrap signal of a single scalar that acts as a discriminator usually contains less information. To alleviate this situation, the Leak-GAN model not only provides discriminator information to the generator, but also applies a hierarchical reinforcement learning architecture to the generator in order to give the leaked information to G_θ.The block diagram of the Leak-GAN model is shown in Figure 2.

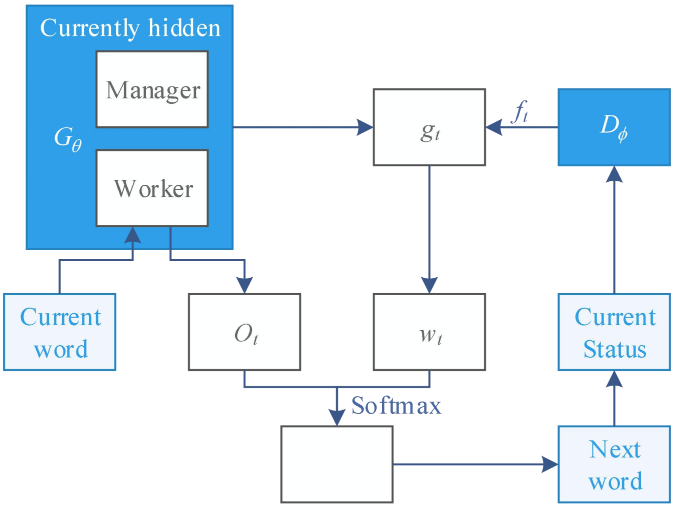

Figure 2. Leak-GAN model framework.

First, the manager and worker modules of the generator are introduced, both of which start from an all-zero hidden state, denoted as h_0^W and h_0^M . In each step, the discriminator sends the leaked feature vector f_t to the manager and generates the target vector g_t in conjunction with its current state:

$$\hat{g}_t, h_t^M = M\left(f_t, h_{t-1}^M; \theta_m\right) \tag{2}$$

$$g_t = \frac{\hat{g}_t}{\left\|\hat{g}_t\right\|} \tag{3}$$

Where $M\left(\cdot;\theta_m\right)$ denotes the LSTM module used by the manager, the current hidden vector is denoted as h_t^W , and the leaked feature vector is f_t.

The discriminator D_ϕ can be expressed as Eq:

$$D_\phi(s) = \text{sigmoid}\left(\phi_f \mathrm{F}\left(s;\phi_f\right)\right) = \text{sigmoid}\left(\phi_f f\right) \tag{4}$$

where $\text{sigmoid}(z) = 1/\left(1+e^{-z}\right)$, $\mathrm{F}\left(.;\phi_f\right)$ denotes the feature extractor of the CNN, and $\mathrm{f} = \mathrm{F}\left(\mathrm{s};\phi_f\right)$ is the feature vector of the last layer in D_ϕ.

Next is the current sum of c-targets generated with the weight matrix W_ψ , where ψ is linearly varying and the target embedding vector w_t is defined as follows:

$$w_t = \psi\left(\sum_{i=1}^{\mathrm{c}} g_{t-i}\right) = W_\psi\left(\sum_{i=1}^{\mathrm{c}} g_{t-i}\right) \tag{5}$$

The current note x of the generated melody is used as input to the worker module. In order to obtain the final spatial distribution of the next note to be selected, a matrix product is realized from O_1 over w_t by softmax at the current state of the manager:

$$O_t, h_t^W = W\left(x_t, h_{t-1}^W; \theta_w\right) \tag{6}$$

$$G_\theta\left(\cdot \mid s_t\right) = \text{softmax}\left(O_t \cdot \mathrm{w}_t / \alpha\right) \tag{7}$$

where h_t^W is the cyclic hidden vector of the LSTM, and $W\left(\cdot;\theta_w\right)$ denotes the worker module. O_t denotes the vector matrix of all notes, and α is a temperature parameter that indicates the generated entropy.

3.4. Experimental results and analysis

In this section, based on the music listening experiments, the designed music automatic generation model Leak-GAN is applied to test the effectiveness of the model.

3.4.1. Music Listening Experiment

Experiments on signal deletion in the frequency domain of a musical melody show the phenomenon of note pitch jitters. These jitters resulted in dissonant semitones, and these semitones would disrupt the original tonality. It has been found that music that is well liked generally has a corresponding tonality, and the destruction of the tonality significantly reduces people's ratings of the music. In this paper, we attempt to explore the effect of frequency-domain variations of musical melodies on the tonality of music in the time domain.

The tonal distribution of a certain music melody after frequency domain sequence length addition and deletion is shown in Fig. 3. From the tonality of the musical melody, it is clear that the tonality of this melody is most likely to be ^{b}D, and almost all the notes are located in this tonality. In other words, almost all the notes conform to the pitch distribution in the key of #C, while all the other keys score lower than #C.

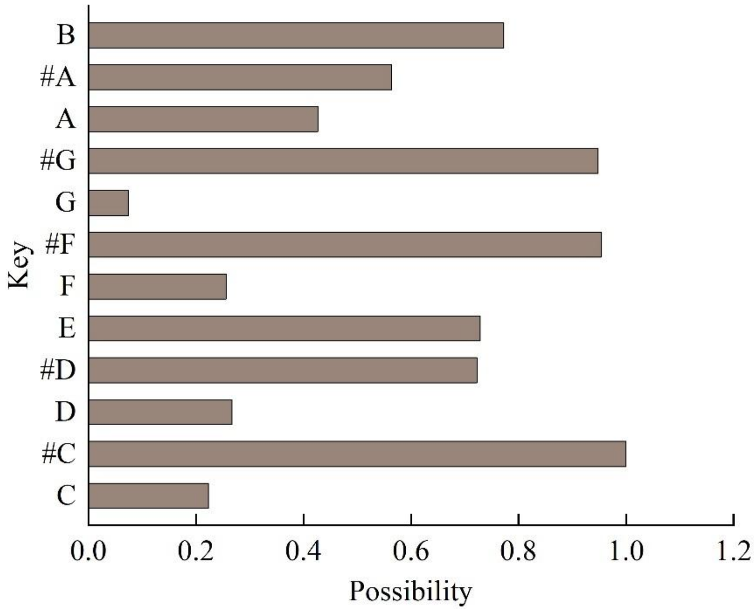

Figure 3. Tonal distribution of a certain musical melody.

The frequency domain scores of the melodic tonality of the passage are shown in Figure 4. It can be seen that there is no significant tonality regardless of the frequency domain signals, which demonstrates the devastating effect of the deletion of high frequency signals in the frequency domain on the melodic tonality.

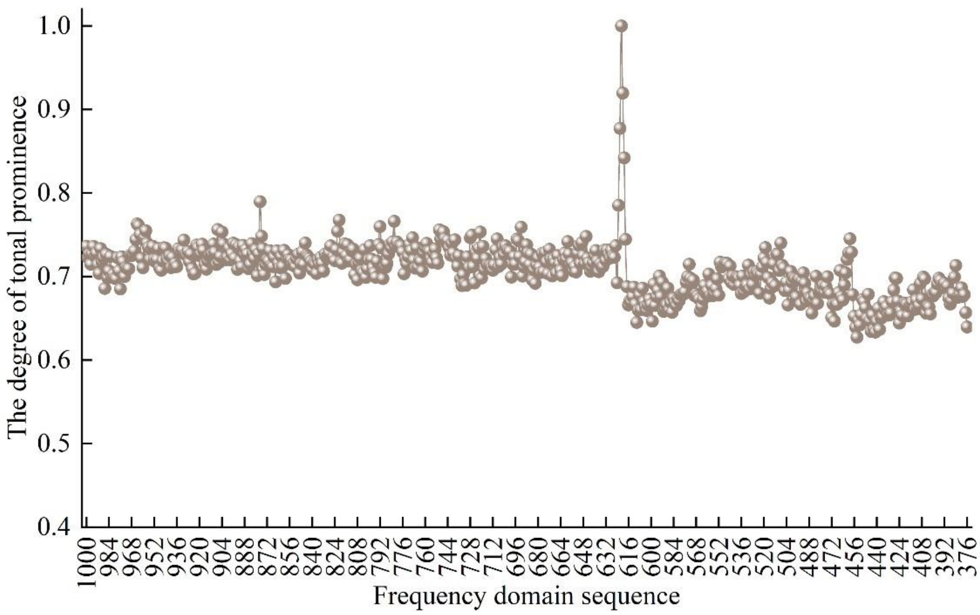

Figure 4. The significant degree of tonal change.

The tonal variation of the musical melody is shown in Fig. 5, where 0~11 of the vertical coordinates represent the keys C, #C, D, #D, E, F, #F, G, #G, A, #A, B. It can be seen that regardless of the change in the length of the frequency domain sequence, the most probable tonality of the musical melody basically maintains in the key of #C, which indicates that it may still be possible to extract the original tonality from the melody after the deletion of the frequency domain sequence and the the extracted tonality can be further optimized.

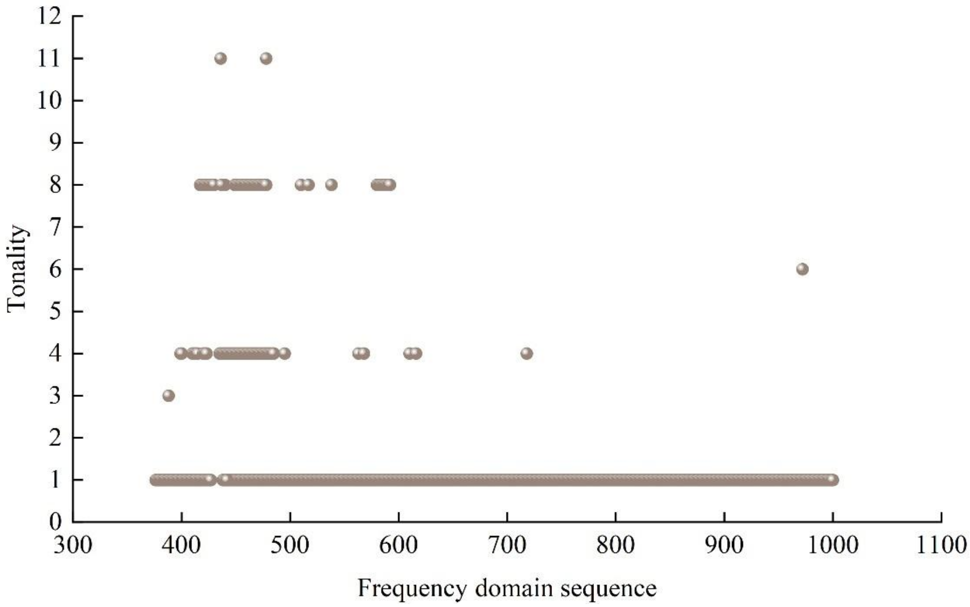

Figure 5. Tone change.

From the above experimental results, it can be seen that when the length of the frequency domain sequence gradually decreases or increases, the dynamic time-regularized distance between the melody transformed back to the time domain and the original melody will gradually increase, and the tonality becomes insignificant.

The trend of the dynamic time-regularized distance between the melody transformed back to the time domain and the original melody in the case of changing the length of the frequency domain sequence is shown in Fig. 6. It can be observed that if the length of the frequency domain sequence of the musical melody is changed a little, the dynamic time-adjusted distance between the melody transformed back to the time domain and the original melody changes greatly. Although the dynamic time-adjusted distance has an obvious decreasing trend near the original frequency domain sequence length, it only drops to zero at the original frequency domain sequence length.

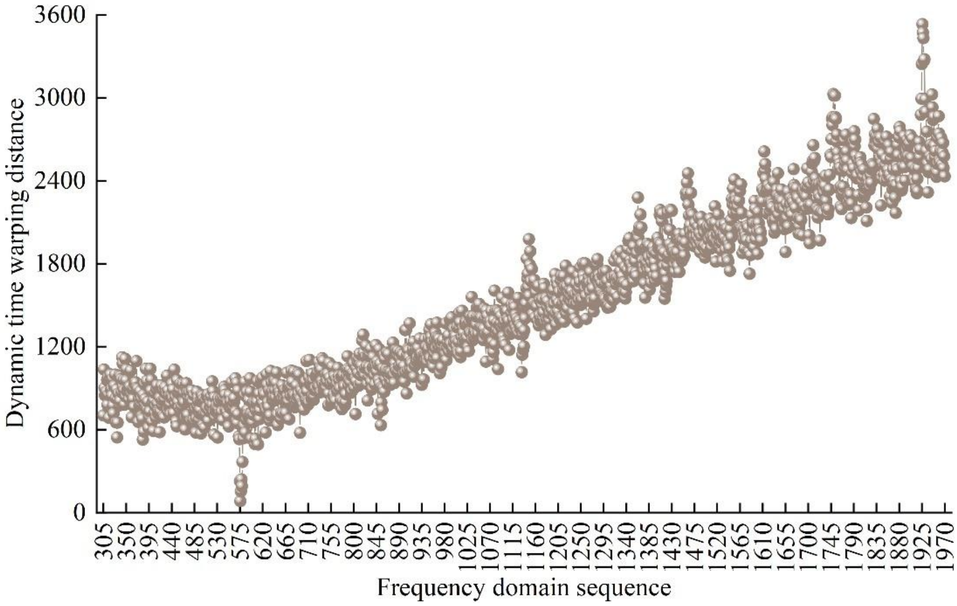

Figure 6. The changing trend of dynamic time warping distance.

Based on the data in Fig. 6, this paper makes the following analysis: the music listening sensation is highly susceptible to changes in the length of frequency domain sequences. Based on this conclusion, this paper conducted the following experimental proof.

Sixty students were selected to participate in this experiment, who came from different fields, such as natural sciences, social sciences and art-related disciplines. The frequency domain sequences of musical melodies were firstly processed with different frequency intensities of interference, then the processed melodies were disrupted, and finally, each student was invited to evaluate the musical listening perception of these melodies. The results of the participants' ratings of musical listening sensation under different frequency intensity interference are shown in Figure 7. It can be seen that as the frequency intensity of interference increases, the rating of music listening perception decreases rapidly, from the beginning of 9.17 points to 3.23 points. However, after the interference frequency intensity exceeded a certain range, the decreasing trend of the ratings became less obvious again, and instead increased slightly by 0.13 points.

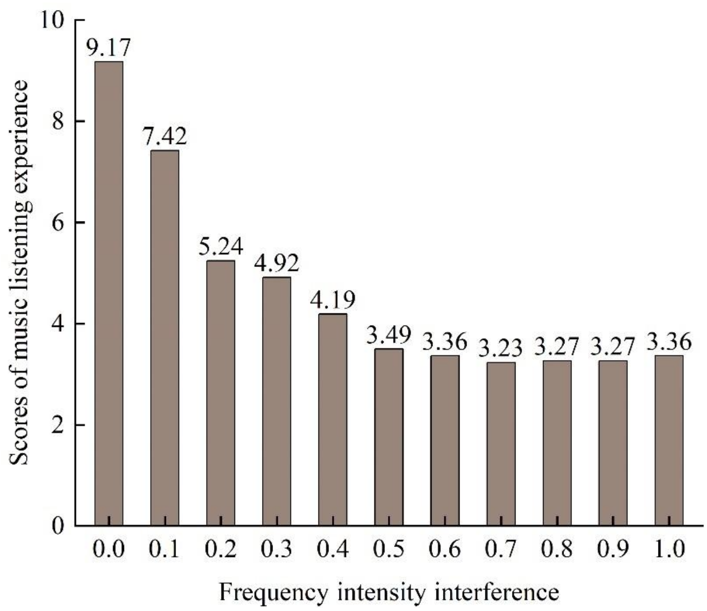

Figure 7. The scoring results of participants' music listening experience.

The following conclusion can be drawn from Fig. 7: In the frequency domain, changes in the frequency domain sequence are transmitted to the notes in the time domain, which in turn affects the musical listening experience. From the nature of the frequency domain transformations, it is known that the pitch and duration of each note in the time domain is determined by the frequency domain sequence. This creates a very big dilemma: it is very difficult to precisely control each note in a melody using frequency domain sequences, and this method seems to be more difficult than generating notes directly in the time domain, because frequency domain sequences have both complex and continuous properties. The number of possible values of the frequency domain sequence signal is greater than the discrete number of note pitches in the time domain for the same sequence length.

3.4.2. Music Generation Modeling Application Experiments

(1) Experimental dataset

This paper is mainly based on the original GAN using pytorch network framework to generate the dataset for melodic use, the music genre belongs to popular music. Number of melodic bars: 51534bars. Data format: midi format. Source: Theory Tab. number of tracks: two tracks, i.e. melody track + chord track.

The experimental dataset consists of three parts, real melody, starter bars and chords. The dataset of real melodies and starter bars is divided into training and test sets in the ratio of 8:2, where the number of bars in the training set is 41,227 bars and the number of bars in the test set is 10,307 bars. The number of chord bars, because in the training and generation as well as each layer of transposition convolution and convolution process, are added to the same part of the chord content, so the number of chord bars is not proportional allocation, he is in the training process, mainly play a role in making the generation of the melody is more stable, in line with the generative properties of music theory. No validation set is set in this experiment.

(2) Experimental Procedure

There is no scientific and rigorous objective evaluation standard in the generation of music melodies, and the main evaluation method is mainly based on the user's subjective evaluation. The evaluation is based on the coherence, pleasantness and interestingness of the generated music melody, and the music melody generated by the baseline model GAN is used as the control group, and the music generation model based on Rank-GAN, the music generation model based on Seq-GAN [20], and the music generation model based on Leak-GAN are used as the control group, and the four models are trained for 300 rounds of training, and the generated music files are then transferred to the validation set for the evaluation. The four groups of models were trained for 300 rounds, and the generated music files were converted from npz format to midi format by pianoroll of python library, and then the music melodies in midi format were converted to MP3 format by MIDI3 pro software, and finally the generated music melodies were evaluated and analyzed. A total of 60 people were evaluated, 48 of whom were general listeners and 12 were music professionals, and the results were evaluated in terms of coherence, pleasantness, and innovativeness of the melodies. The evaluation method was based on a five-point scoring system, with 1 being the worst effect, 5 being the best effect, and the rest of the scores were similar.

(3) Experimental results and analysis

Through the weighted average of 60 people's scoring, the following experimental subjective evaluation results were obtained as shown in Table 1. It can be seen that the music melody generated by the Leak-GAN model in this paper in the three aspects of coherence, pleasantness and innovativeness of the general audience score of 4.1, 4.1, 3.9, respectively, and the professionals scored 3.9, 4.1, 3.5, are better than other comparative models, which verifies the practicality of the model built.

Table 1. Subjective evaluation of model experiment results.

Model	General audience			Music professional		
	Coherence	Pleasantness	Innovation	Coherence	Pleasantness	Innovation
GAN	3.2	3.3	3.4	2.4	2.7	2.3
Rank-GAN	3.3	3.3	3.3	2.8	2.8	3.4
Seq-GAN	3.5	3.8	3.8	3.2	3.2	2.8
Leak-GAN	4.1	4.1	3.9	3.9	4.1	3.5

In the two-track music generation based on GAN network, the evaluation results were analyzed using the weighted average method. For the results generated by the four sets of models, the evaluation results of 48 general listeners were calculated at 40% and the evaluation results of 12 music professionals were calculated at 60% as in Equation (8):

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_k f_k}{\sum_1^k f_i} \tag{8}$$

For the three kinds of performance coherence, pleasantness and innovation of the weight of the evaluation indexes according to 5:3:2 to analyze, resulting in four groups of model performance analysis results as shown in Table 2. According to the analysis results, it can be concluded that the performance of the four models gradually improves, and the music melody generated by the Leak-GAN model in this paper is more realistic and pleasing to the ear, with a total score of 3.95.

Table 2. Weighted average results of musical melodies generated by four groups of models.

Model	General audience	Music professional	Total score
GAN	3.27	2.47	2.79
Rank-GAN	3.30	2.92	3.07
Seq-GAN	3.65	3.12	3.33
Leak-GAN	4.06	3.88	3.95

In music, the core of the generated results is the pleasantness of the generated music, so the pleasantness of the music melodies generated by the four groups of models were correlated with coherence and innovation respectively, according to the Pearson correlation coefficients as in Equation (9):

$$r = \frac{N\sum x_i y_i - \sum x_i y_i}{\sqrt{N\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{N\sum y_i^2 - \left(\sum y_i\right)^2}} \tag{9}$$

The results of the correlation analysis are shown in Table 3. It can be concluded that the pleasantness of the musical melodies generated by all four groups of models has a significant positive correlation with the coherence and innovation of the music, among which the correlation of the musical melodies generated by the Leak-GAN model in this paper is even higher, reaching 0.797 and 0.529, respectively.

Table 3. Results of correlation analysis.

Model	Correlation coefficient between pleasantness and coherence	Correlation coefficient between pleasantness and innovativeness
GAN	0.685	0.441
Rank-GAN	0.706	0.398
Seq-GAN	0.584	0.484
Leak-GAN	0.797	0.529

Further synthesizing Tables 1 and 2 shows that the Leak-GAN-based music generation model outperforms the remaining three models in terms of coherence, pleasantness and innovativeness. The three proprietary models outperform the baseline model regarding pleasantness and coherence. The general public's evaluation of the melodies generated by the four sets of models tends to increase gradually. In summary, the comprehensive performance of the Leak-GAN-based music generation model has been greatly improved, optimized and innovated based on the baseline model.

4. Experiments in Exploring Students' Aesthetic Experience in Music Education

In this chapter, music education experience was used as the independent variable, and two groups of subjects with many years of music training experience and no relevant experience at all were selected to explore the effects of music emotion type, music preference and music training experience on students' aesthetic experience as a means of providing suggestions and references to the issue of enhancing students' aesthetic experience.

4.1. Experimental methodology

4.1.1. Subjects

In this study, the 160 college student subjects (80 each from music training and non-music training, mean age 20.35 years, standard deviation 1.19) were divided into a high preference group (48) and a low preference group (48) to examine the differences in music aesthetic experience among college students, with the 30% before and after music preference scores as the grouping principle. All subjects were right-handed, had normal hearing and speech, normal naked or corrected vision, and were free of somatic diseases and mental disorders. Prior to the experiment, they were asked to avoid substances or drugs that might affect attention.

4.1.2. Experimental design

A three-factor mixed experimental design was adopted, with 2 (music preference: high preference group, low preference group) ×3 (music emotion type: happy, calm, sad) ×2 (music training: expert group, novice group). Music emotion type was the intra-subject variable, and music preference and music experience were the inter-subject variables. The dependent variable is the degree of preference (1 "very much like", 6 "very little like"), aesthetic degree (1 "not beautiful", 9 "very beautiful"), arousal degree (1 "very calm", 9 "very excited"), and familiarity degree (1 "very familiar", 6 "very unfamiliar") of the music evaluated by the subjects based on their personal experiences.

4.1.3. Experimental materials

45 music clips that express happiness, calmness and sadness 1/3 each. Includes opera, symphony, folk music, and other art forms. The repertoire was drawn from materials used in previous studies, relevant research findings, and ratings by music professionals. Quality audio of musical performances was selected and edited into approximately 30-second recorded segments to ensure a sense of musical integrity and to control for level of expression and familiarity. All clips were rated by 100 university students as effective in conveying a specific mood. Audio bitrate 64kbps, sound quality sampling frequency 44Hz, two-channel stereo.

4.1.4. Experimental procedures

The experiment was conducted in a standardized behavioral laboratory with good multimedia and acoustics of the equipment, where the subjects could first turn the volume to the right level. E-Prime 2.0 was used for programming and execution. The emotion ratings include what kind of emotion this musical performance is trying to express and how intense this emotion is, and what kind of emotion this performance makes you really experience and how intense this emotion is. To avoid making the choices too easy, the emotion type rating retains nine options of wonder, transcendence, lyricism, nostalgia, calmness, strength, delight, tension, sadness, and fear, with a 9-point intensity rating. The first two questions point to musical emotion recognition and the last two to musical emotion experience. The order of the four ratings remained unchanged because the order did not significantly affect the results and to avoid misjudgments caused by changing the order. Ratings of correctness and intensity were recorded, as well as reaction times for the first rating item.

4.2. Experimental results and analysis

The results of the post-hoc t-tests and within-group post-hoc t-tests for experts' and novices' music aesthetic experience are shown in Tables 4 and 5, respectively.

Repeated measures ANOVA showed a significant main effect of musical mood type on musical aesthetics. Post hoc t-tests revealed that experts in both groups were significantly higher than novices in music aesthetics scores. The within-group difference was significant for the low preference group, with experts scoring higher on musical aesthetics than novices. The interaction between musical emotion type and group was not significant ($p>0.05$). In terms of emotional experience, the low preference group had a significant main effect on musical mood type, and post hoc t-tests found that experts in both groups scored significantly higher than novices on emotional experience, and the main effect of group and the interaction between musical mood type and group were not significant ($p>0.05$). The main effect of musical mood type was significant for music-induced arousal, and post hoc t-tests found that both groups of experts scored significantly higher than novices on musical arousal, and the main effect of group as well as the interaction effect of musical mood type and group were not significant ($p>0.05$).

Table 4. Post-event test results of music aesthetic experiences of experts and novices.

Variable		High preference group (Newbie + Experts)			Low preference group (Newbie + Experts)		
		F (df)	*p*	η^2	*F* (df)	*p*	η^2
Aesthetic sense	Music emotion type	12.83 (2, 45)	0.000***	0.24	40.24 (2, 45)	0.000***	0.54
	Group	0.019	0.93	0.00	4.83(2, 45)	0.03*	0.13
	Music emotion type * Group	0.20	0.69	0.006	2.01 (2, 45)	0.15	0.06
Experience	Music emotion type	1.93 (2, 45)	0.18	0.06	4.85 (2, 45)	0.01**	0.12
	Group	0.04 (2, 45)	0.88	0.001	0.61 (2, 45)	0.46	0.02
	Music emotion type * Group	0.57 (2, 45)	0.84	0.001	0.009 (2, 45)	0.927	0.00
Wake up	Music emotion type	39.21 (2, 45)	0.000***	0.47	93.51 (2, 45)	0.000***	0.72
	Group	3.04 (2, 45)	0.12	0.09	1.62 (2, 45)	0.24	0.05
	Music emotion type * Group	0.62 (2, 45)	0.451	0.03	0.21 (2, 45)	0.658	0.02

Table 5. The intra-group post-event test results of the music aesthetic experiences.

Variable	High preference Group (M±SD)			Low preference Group (M±SD)		
	Aesthetics experience	Novice (N=22)	Expert (N=26)	Aesthetics experience	Novice (N=27)	Expert (N=21)
Sad music	Aesthetic sense	6.51 (0.98)	6.33 (0.89)	Aesthetic sense	4.31 (1.42)	4.86 (1.03)
	Experience	2.40 (0.91)	2.36 (0.54)	Experience	2.05 (0.61)	2.05 (0.59)
	Wake up	4.15 (1.47)	4.57 (1.61)	Wake up	3.54 (1.31)	3.71 (1.12)
	Familiar	3.39 (0.92)	3.15 (0.83)	Familiar	4.08 (0.61)	3.98 (0.39)
Calm music	Aesthetic sense	7.08 (1.11)	6.62 (0.96)	Aesthetic sense	5.69 (1.27)	5.94 (1.21)
	Experience	2.59 (0.97)	2.43 (0.69)	Experience	2.21 (0.52)	2.31 (0.73)
	Wake up	3.48 (1.49)	3.95 (1.44)	Wake up	3.35 (1.13)	3.42 (1.31)
	Familiar	3.43 (0.97)	3.18 (0.75)	Familiar*	3.82 (0.65)	3.42 (0.54)
Happy Music	Aesthetic sense	6.04 (1.16)	6.36 (1.09)	Aesthetic sense***	4.82 (1.31)	6.25 (1.01)
	Experience	2.41 (1.04)	2.52 (0.93)	Experience	2.03 (0.64)	2.19 (0.77)
	Wake up*	5.25 (1.35)	6.24 (1.05)	Wake up*	5.35 (1.26)	6.03 (0.94)
	Familiar***	3.34 (0.95)	2.90 (0.86)	Familiar**	3.39 (0.87)	3.08 (0.61)

5. Conclusion

This paper explores the inheritance and innovation application strategy of AI technology in traditional music education, constructs a music automatic generation model based on Leak-GAN, and carries out experimental investigation on the enhancement of students' aesthetic experience in music education. The conclusions obtained are as follows:

In music teaching, teachers can effectively integrate artificial intelligence technology by encouraging students to create rhythms and refining the content and methods of improvisation and choreography teaching, which can be used to realize the inheritance and innovation of music knowledge and skills and enhance students' aesthetic experience.

Music listening is highly susceptible to changes in the length of frequency domain sequences. With the enhancement of the interference frequency intensity, the score of music listening sense will decrease rapidly, but after the interference frequency intensity is beyond a certain range, the decreasing trend of the score becomes insignificant again. Meanwhile, in the frequency domain, the changes in the frequency domain sequence are transmitted to the notes in the time domain, which in turn affects the music listening perception.

The performance of the four models, GAN, Rank-GAN, Seq-GAN and Leak-GAN, in automatically generating music melodies gradually improves, among which the music melodies generated by the Leak-GAN model in this paper are more realistic and pleasing to the ear, with a total score of 3.95, and the pleasingness of the generated music melodies has a higher correlation with the coherence and innovativeness of the music, which reaches 0.797 and 0.529.

In addition, the main effect of musical mood type was significant in the musical aesthetic experience dimensions of musical aesthetics, emotional experience, and music-induced arousal. Experts in both the high and low preference groups had higher aesthetic experience scores than novices.

About the Author

Ning Li (born August 1978), Han Chinese, from Zhengzhou, Henan Province, ID number: 410103197808103225, holds a master's degree. She is a lecturer at the School of Education Science, specializing in art education, music education, dance education, and dance performance.

References

[1] Yuan, Z. (2025). The Relationship between Chinese Music History and The Study of Traditional Chinese Music. Mediterranean Archaeology & Archaeometry, 25(1).

[2] Hu, M. (2022). Features of singing in Chinese pop and traditional music: The influence of the music genre on vocal music. Musica Hodie, 22.

[3] Xu, N. (2018). Analysis of the Correlation Between Folk Music Education and Chinese Traditional Culture. Educational Sciences: Theory & Practice, 18(5).

[4] Zhou, Y., & Yu, F. (2024). Integration of traditional Chinese music: an evaluation of the interactive influence between traditional music and aesthetic thought. Trans/Form/Ação, 47(5), e02400180.

[5] Tu, X. (2022). Opportunities and Challenges of Chinese Music Curriculum Standards: The gap between urban and rural areas music education. Unpublished master’s thesis, University of Auckland.

[6] Huanyuan, Z. (2022). Problems in China's college music teaching in recent years. International Journal of Management and Education in Human Development, 2(02), 458-460.

[7] Wu, Y. B., Lin, H., & Zhu, W. H. (2021, June). Function design of music online education network virtual classroom platform. In International Conference on E-Learning, E-Education, and Online Training (pp. 631-643). Cham: Springer International Publishing.

[8] Ruthmann, S. A., & Hebert, D. G. (2018). Music learning and new media in virtual and online environments. Creativities, technologies and media in music learning and teaching: an Oxford handbook of music education, 254-272.

[9] Yan, J., & Xia, X. (2024). Interactive audio-visual course teaching of music education based on VR and AI support. International Journal of Human–Computer Interaction, 40(13), 3552-3559.

[10] Ng, D. T., Ng, E. H., & Chu, S. K. (2022). Engaging students in creative music making with musical instrument application in an online flipped classroom. Education and information Technologies, 27(1), 45-64.

[11] Xiao, H. (2025, April). Personalized Learning Path Planning and Effect Verification of Music Education Information System Empowered By Artificial Intelligence. In Proceedings of the 2025 International Conference on Artificial Intelligence and Educational Systems (pp. 443-448).

[12] Lin, F., & Chen, F. (2025). Intelligent music education system: utilizing algorithms for personalized learning experience. International Journal of High Speed Electronics and Systems, 2540253.

[13] Young, G. W., Murphy, D., & Weeter, J. (2018). A functional analysis of haptic feedback in digital musical instrument interactions. In Musical Haptics (pp. 95-122). Cham: Springer International Publishing.

[14] Ma, Y., & Chen, Y. (2024). Exploring the model of contemporary Chinese ethnic musical instrument improvement mechanisms: Based on grounded theory. Sage Open, 14(1), 21582440241235018.

[15] Hwang, G. H., Chen, B., & Sung, C. W. (2019). Impacts of flipped classrooms with peer assessment on students’ effectiveness of playing musical instruments–taking amateur erhu learners as an example. Interactive Learning Environments, 27(8), 1047-1061.
[16] Sysoieva, S., Ovcharenko, N., & Chebotarenko, O. (2021). Future music and art educators' professional development: theoretical and technological issues. Ukrainian Journal of Educational Studies and Information Technology, 9(3), 18-33.
[17] Zdravić-Mihailović, D. (2020). The role of a teacher in music education of young professional musicians. Facta Universitatis, Series: Teaching, Learning and Teacher Education, 057-066.
[18] Burt-Perkins, R., & Lebler, D. (2008). ‘Music isn’t one island’: The balance between depth and breadth for music students in higher education. In Educating Musicians for a Lifetime of Learning. International Society of Music Education (ISME).
[19] Hu Jiaxin, Ge Zhaohui & Wang Xiaohua. (2022). The Psychological Education Strategy of Music Generation and Creation by Generative Confrontation Network under Deep Learning. Computational intelligence and neuroscience,2022,3847415-3847415.https://doi.org/10.1155/2022/3847415.
[20] Jiafa Zhang, Hong Zou, Zifeng Zeng, Weijie Xu & Jiawei Jiang. (2024). Feasibility of Using Seq-GAN Model in Vulnerability Detection of Industrial Control Protocols. Journal of Cyber Security and Mobility,13(3),393-416. https://doi.org/10.13052/JCSM2245-1439.1333.