International Journal for Housing Science and Its Applications
Publish September 11, 2025. Volume 47, Issue 1 Pages 827-838

https://doi.org/10.70517/ijhsa47172

The Impact of Artificial Intelligence on Creative Thinking in Music Composition (Teaching)

Qi Liu[1,*]
[1] Department of Art and Technology, Zhejiang School of Music and Dance, Communication University, Hangzhou, Zhejiang, 310018, China
Corresponding authors: (e-mail: 20140048@cuz.edu.cn).

Abstract With the continuous progress of artificial intelligence algorithms, music creation has gradually developed from purely manual operation to intelligent assistance, a process that not only optimizes the efficiency of music creation, but also opens up new paths of creative thinking. This study explores the impact of artificial intelligence on creative thinking in music composition and teaching. By introducing AI-based music composition models and teaching methods, its performance in practical applications was analyzed. The study used scoring data from 30 professional listeners and 25 general listeners and compared them with the AI scoring system. The results show that the difference between the AI ratings and the ratings of the professional listeners is about 1.5 points, indicating that AI can more accurately simulate the perception of professional listeners in music composition evaluation. In addition, the model training generated the highest music scores at an iteration number of 60k, which were 79.249 points for the artificial rating and 80 points for the model rating, respectively. The study shows that AI has a significant auxiliary role in music creation and teaching, which can promote the diversity of music creation and personalization of teaching, and enhance students' understanding and interest in music.

Index Terms Artificial intelligence, music composition, music teaching, scoring model, personalized learning, music evaluation

I. Introduction

In today's highly technologically advanced era, artificial intelligence (AI) has played a great role in various fields, especially in the teaching of music composition [1]. Music has always been an integral part of human life, and the intervention of AI has further promoted the development of music composition [2].

AI is applied to the field of music composition in various ways. Among them, the most common is the algorithmic model based on machine learning, through the training of a large number of musical works, the machine learning habit algorithm can learn and imitate the human music composition style and generate new musical works [3]-[5]. In this process, AI is able to automatically recognize and extract features of music, such as melody, chords, and rhythm, and then create new music based on these features.AI is also able to create music based on a specific theme or emotion [6]-[8]. For example, by inputting a textual description, AI can generate a corresponding musical composition based on the emotion of the text. This kind of work makes music creation more proactive and innovative. In addition to machine learning, AI has also been applied to the development of music creation tools [9], [10]. For example, some music production software integrates AI algorithms that can automatically modify and improve the user's music creation set process, providing creative inspiration and technical support [11], [12]. This enables music creators to express their ideas more efficiently and quickly realize their music creation goals [13]. While AI performs many advantages in music creation, including liberating the creative ability of music creators, more improving the efficiency of music creation, etc., but at the same time, there are also shortcomings such as lack of depth, emotion and subtlety, so it is important to correctly see the advantages and disadvantages of AI in music creation [14]-[17].

The core of this study is to explore how artificial intelligence promotes creative thinking in music composition and its effectiveness in music teaching. The research methodology includes comparing and analyzing the differences between AI and human evaluation, assessing the accuracy of AI scoring models; analyzing the effects of different training iterations on musical compositions through experimental data, as well as the role of AI in music teaching in stimulating students' learning efficiency and interest. Specifically, firstly, by comparing the results of AI scoring and human scoring in music composition, we analyze the advantages and shortcomings in evaluation; secondly, by tracking the training process of the model, we assess the quality enhancement of music composition with different number of training iterations; and lastly, through the teaching experiments, we assess the potential of AI to improve students' musical understanding and interest in composition.

II. AI-based music creation and teaching

II. A. The Role of Artificial Intelligence in Music Education

II. A. 1) Personalized Learning Experience Enhancement

Firstly, accurate resource delivery will be realized: by taking into account students' academic progress, knowledge mastery level and personal interests, Knowledge Graph can accurately recommend appropriate learning materials, such as textbooks, videos and exercises, to avoid aimless searching in the learning process and significantly improve learning efficiency. Secondly, the planning of personalized learning paths will be strengthened: by analyzing students' existing knowledge structure and learning goals, learning paths can be customized to meet individual characteristics, guiding students to learn in the most suitable way for themselves, and helping them to build up a systematic knowledge system, deeply understand and master what they have learned, thus realizing the personalization of education. Finally, the organization of teaching content and curriculum design will be optimized: the use of knowledge mapping makes the hierarchical structure and logical relationship of subject knowledge visualized, which is easy for teachers and students to grasp, so as to build a complete knowledge system.

II. A. 2) Enrichment and expansion of teaching resources

AIGC technology is capable of generating a wide range of content types, including but not limited to text-based lyrics, scripts, poems, etc. On the audio side, it is able to cover areas such as music composition and speech synthesis. For music education resources, AIGC technology firstly realizes a revolutionary improvement in content creation efficiency, enabling rapid generation of preliminary drafts. Secondly, it promotes the democratization of content creation, lowers the threshold of creation, and enables ordinary individuals to participate in content creation. At the same time, AIGC technology stimulates creative inspiration, and creators can be inspired by the generated content to improve their own works; AIGC technology also expands the diversity of content, and is able to generate a variety of styles of content that meet the needs of different audiences.

II. A. 3) Creative assistance and innovation stimulation

Large-scale modeling will play a key supporting role in the process of music composition. Students and teachers can utilize the large-scale model to generate and combine musical elements efficiently, for example, by inputting information such as keywords, to quickly obtain music clips or works of a specific style. Modification and refinement on this basis will help cultivate innovative thinking and creative ability, as well as enhance creative interest and self-confidence.

However, there are still some problems with the current AI music generation technology. In terms of sound quality, due to the complexity of music signals, the limitations of arithmetic power and the varying quality of training data, the sound quality of AI-generated music is unsatisfactory, for example, the reproduction of audio details is not accurate enough, and there are fuzzy distortions and other problems.

II. A. 4) Enhancing the objectivity and accuracy of teaching evaluation

The application of artificial intelligence-based music information retrieval technology in teaching and learning evaluation has become more and more important, which analyzes the music content in depth through a variety of technical means, thus providing powerful support for teaching. First of all, music content analysis includes melody extraction, rhythm analysis, harmony analysis and timbre recognition. Melody extraction can help students better understand and sing music, and can also be used to retrieve specific melodic fragments. Rhythm analysis focuses on beat tracking and recognition of rhythmic patterns, which is important for the development of automatic accompaniment systems. Harmony analysis helps to understand the structure and emotional expression of the music, which is very useful for teaching harmony. Tone recognition technology is also widely used in the fields of music production and music retrieval.

II. B. Application of Artificial Intelligence in Music Teaching

II. B. 1) Voice recognition

Sound recognition technology, also known as speech recognition technology, is a technology that allows computers and other devices to parse and understand human speech through the use of various algorithms. The core of this technology lies in converting sound signals into digital signals that can be processed by machines, and then analyzing these digital signals through complex algorithms to convert them into text or execute the appropriate commands. The process of voice recognition involves sound acquisition, feature extraction, pattern matching, and final interpretation and execution, a sequence of operations that relies on theories and technologies from a variety of subject areas such as digital signal processing, artificial intelligence (AI), machine learning, deep learning, and so on.

II. B. 2) Tone detection

Pitch detection technique is an audio signal processing technique used to identify and measure pitch information in an audio signal. The technique mainly contains the following five steps:

Step 1: Audio signal acquisition

This operation is the initial stage of pitch detection, which involves the use of a suitable hardware device, such as a microphone or audio interface, to capture the sound signal and convert it into a numerical model for subsequent data processing.

Step 2: Audio Preprocessing

Before the audio signal enters the pitch detection algorithm, pre-processing, such as noise reduction, noise removal, sample rate adjustment, etc., is usually required to improve the accuracy of pitch detection.

Step 3: Feature Extraction

This step involves extracting pitch-related features from the preprocessed audio signal. This usually involves the analysis of time and frequency domain features, such as Short Time Fourier Transform (STFT) or Mel Frequency Cepstrum Coefficients (MFCC).

Step 4: Pitch Estimation

After extracting the relevant features, the pitch estimation algorithm estimates the pitch of the audio signal based on these features. This usually involves the application of pattern recognition or machine learning algorithms.

Step 5: Result Output

Finally, the pitch detection system outputs the result of the pitch estimation, usually presented in notes, hertz (Hz), or other pitch representations. Information, widely used in music analysis, speech recognition, instrument tuning and other fields.

In music teaching, the application of artificial intelligence is gradually penetrating into various fields. Among them, the application of pitch detection technology in music teaching is particularly prominent. Pitch detection technology is a technology that can identify and analyze audio signals, which can help teachers and students understand and master the concept of pitch more accurately, thus improving the effect of teaching.

II. B. 3) Rhythm analysis

Fig. 1 shows the music beat detection method.Rhythm analysis is an audio processing technique that focuses on automatically recognizing and extracting rhythmic patterns from audio signals. It is based on a series of complex algorithms and analysis processes to accurately understand rhythmic changes in music or other sound signals. First through signal acquisition and processing: the technology first acquires the sound signal, followed by filtering, noise reduction and other pre-processing means to optimize the quality of the signal and highlight rhythmic features. Then feature extraction: the algorithm will extract features closely related to the rhythm, such as intensity, frequency and duration, from the processed signal to form a feature set. After that, pattern matching and recognition: the extracted features are matched and compared with known rhythmic patterns to recognize the rhythm in the audio through pattern recognition algorithms, such as Hidden Markov Models (HMMs) or Deep Learning Networks. Final result output: the recognized rhythmic patterns are output, including information such as beat tempo, beat number, and possible music style classification.

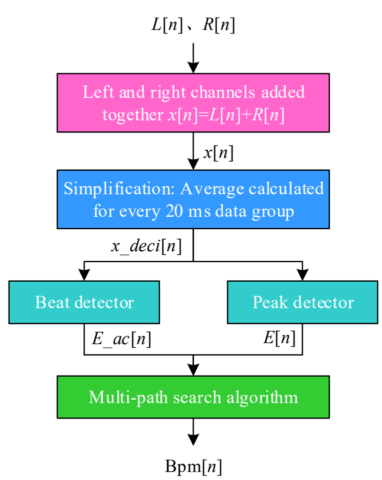


Figure 1: The music festival is a method of testing

II. C. Audio feature extraction representation

II. C. 1) Amplitude spectrum characterization

Spectral transformation refers to the process of converting an audio signal from the time domain to the frequency domain representation, by performing a short-time Fourier transform (STFT) on the signal, and multiplying a time-limited window function $h(t)$ before the signal is Fourier transformed and assuming that the non-smooth signal is smooth within a short time interval in the analysis window, and analyzing the signal segment-by-segment by the shift of the window function $h(t)$ on the time axis A set of localized spectra of the signal is obtained. The STFT of the signal $x(t)$ is defined as:

$$STFT(t,f)=\int_{-\infty}^{\infty}x(\tau)h(\tau-t)e^{-j2\pi f\tau}d\tau \tag{1}$$

In the above equation $x(\tau)$ represents the sampling of the signal in the time domain and $h(\tau-t)$ is the window function.

After the above transformation, the speech signal is represented in complex form, where the real and imaginary part of the spectrum together form the spectrogram. The spectrogram shows the intensity distribution of the signal at different frequencies, presenting the characteristics of the signal in the frequency domain. The amplitude spectrum can be obtained by taking the absolute value of the spectrogram after Fourier transform, which mainly shows the amplitude or intensity of the frequency.

The amplitude spectrum is the most commonly used audio feature for speech signal processing, and with the development of the field of music information retrieval, the amplitude spectrum is also applied to the melody extraction task. However, for polyphonic music, the audio contains multiple sources and harmonic structures, and using amplitude spectrum features for training, the model may face the challenge of recognizing multiple tones and separating different notes. Second, the magnitude spectrum is more sensitive to noise, and when noise is present in the audio, the noise may have an effect on the magnitude spectrum, making the melody extraction subject to noise interference.

II. C. 2) Constant Q-transform characterization

All tones in music are made up of a number of octaves in twelve equal temperament. In twelve equal temperament, an octave is evenly divided into twelve semitones, ensuring that the pitch relationship between each semitone remains fixed. Specifically, in two octaves of the same scale, the frequency of the higher octave is two times the frequency of the lower octave. Thus, in music, the pitch relationship between notes follows an exponential law. In line with this, the constant Q transform (CQT) is a time-frequency transform algorithm with an exponential distribution law. This algorithm is more suitable for processing music signals, so many researchers have applied it to music analysis tasks such as melody extraction.

II. C. 3) Combined Frequency and Periodicity Representation Characterization

The pitch information of polyphonic music is rich and complex, and audio signals produced by different instruments or sound sources may have similar fundamental frequencies and overtone structures. When the fundamental frequencies of two or more tones are very close to each other, their overtones may interact with each other, leading to interference with the stability and accuracy of the pitch detection or analysis process. Combined Frequency and Periodicity Representation (CFP) acquires pitch by examining the consistency between harmonics in the frequency domain and subharmonics in the cepstrum domain, a method that skillfully integrates the complementary advantages of the two feature domains in different frequency ranges and improves the robustness of the pitch detection function against overtones interfering with simultaneous pitches.

The following is an example of a segment of a time-domain music signal x to describe the specific composition of CFP features. Assuming that x is a discrete signal, the time domain representation of this music signal consists of N discrete sample points. Considering each sample point as a vector, denote x as $x=(x_1,x_2,\cdots,x_n)$, which simplifies to $x:=x[n]$, with n being the time index. The spectral representation is obtained by short-time Fourier transform of x, setting the amplitude part as X, and given the N-point DFT matrix F, the high-pass filters W_f and W_t, and the activation function σ_i, the transforms result in the following three data representations:

$$Z_0[k,n]:=\sigma_0(W_fX) \tag{2}$$

$$Z_1[q,n]:=\sigma_1(W_tF^{-1}Z_0) \tag{3}$$

$$Z_2[k,n]:=\sigma_2(W_fFZ_1) \tag{4}$$

The above three formulas represent three traditional pitch saliency functions: Z_0 for the power scale spectrum, Z_1 for the generalized cepstrum (GC), and Z_2 for the generalized spectral cepstrum (GCoS). The index k in Z_0 and Z_2 denotes the frequency, while the index q in Z_1 is the cepstrum. The cepstrum is denoted as the temporal information in the cepstrum spectrum, which focuses on describing the periodicity of the audio signal, which has the same units as time. The nonlinear activation function σ is defined as:

$$\sigma_i(Z) = \left|\mathrm{Re}\,LU(Z)\right|^{\gamma_i}, i = 0,1,2 \tag{5}$$

where $0 < \gamma_i \leq 1$, $\mathrm{Re}\,LU$ is the rectified linear unit activation function, and $|\cdot|^{\gamma_i}$ is the root function at the element level, which in this paper is set to $(\gamma_0, \gamma_1, \gamma_2) = (0.24, 0.6, 0.1)$. W_f and W_t are two high-pass filters designed on the basis of diagonal matrices with cutoff and cutoff cepstrum frequencies k_c and q_c, respectively, denoted as:

$$W_{f\,or\,t}[l,l] = \begin{cases} 1, & l > k_c \ or \ q_c \\ 0, & otherwise \end{cases} \tag{6}$$

The W_f and W_t filters can be used to remove the slowly varying part of the spectrum, making the frequency variation fluctuations more pronounced. After obtaining the above three feature representations, the Z_0, Z_1 and Z_2 representations need to be mapped to the logarithmic frequency scale in order to better fit the perceptual scale of pitch. Therefore, two filter banks need to be set up: a log-transformed filter and a triangular filter, and the filter parameters can be set according to requirements. The Z_0, Z_1 and Z_2 are first converted to logarithmic frequencies, and then the triangular filter is used to determine the frequency range, and finally the standard CFP representation is obtained.

II. D.Audio Recognition Fusion Model Design

II. D. 1) Deep Neural Network Algorithm

Multilayer perceptron, as the name suggests, is a hierarchical structure with certain rules formed by superimposing multiple single-layer perceptrons through front and back concatenation, and multilayer perceptron can be regarded as the prototype of neural network in current deep learning. A single perceptron consists of two layers of neurons as shown in Figure 2.

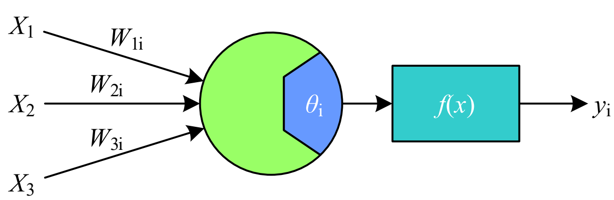


Figure 2: Perceptron structure

$$y_i = \mathrm{sgn}\left(\sum_{j=1}^{m} w_{ji} x_i + \theta_i\right) = \begin{cases} 1, & if \ \sum_{j=1}^{m} w_{ji} x_j + \theta_i \geq 0 \\ 0, & if \ \sum_{j=1}^{m} w_{ji} x_j + \theta_i < 0 \end{cases} \tag{7}$$

The output signal y_i of the output layer is obtained from the input signal x_i received by the input layer from the outside world through the internal processing of the perceptron, and the M-P neuron as the "threshold logic unit" is the output layer. For a certain data set, the samples are selected as inputs to the perceptron network, and the perceptron will automatically adjust the weights w_i and thresholds θ_i in each round of iterative learning process, so that the w_i and thresholds θ_i will be close to the infinite tendency, so as to fully understand the deeper features of the data samples. After each round of learning, the result is multiplied with the learning rate α by making a difference with the result of the previous round, and the result is used as the magnitude of the change of w_i, which is continuously iterated and gradually optimized. The formula for Δw_i is as follows ($\alpha \in (0,1)$ denotes the learning rate):

$$\Delta w_i = \alpha(y - y^*) \quad w_i \to w_i + \Delta w_i \tag{8}$$

The output of each layer is fully connected to the input of the next layer, and the output of neurons in the same layer cannot be connected to each other, nor can it be directly connected to the neurons behind it across one or more neurons, so that the simplest neural network structure is formed, which is generally referred to as a "multi-layer feed-forward neural network". The neurons in the input layer of the first perceptron only receive external inputs, and the dimension of the external input data determines the dimension of the input data of the first perceptron. The first perceptron is not connected to other outputs, the middle hidden layer and the output layer process the data of the input samples, and the judgment of the target category at the end is output by the last neuron.

II. D. 2) Gradient descent optimizer

Since the number of parameters of deep neural network models is often very large, often with millions or more parameters, it is necessary to use an optimization algorithm based on gradient descent to reasonably optimize these model parameters in order to improve the stability of the model and the ability to resist overfitting.

(1) Adaptive Gradient Algorithm: AdaGrad is an optimization algorithm with adaptive learning rate, which can dynamically adjust the learning rate according to the gradient history information of each parameter. Its main idea is to adaptively reduce the learning rate of the parameters with small gradients that occur frequently, and adaptively increase the learning rate of the parameters with large gradients that occur frequently, in order to achieve a better convergence effect.

Specifically, for each parameter θ_i to be optimized, AdaGrad maintains a historical accumulation of the sum of squares of the gradients G_i and uses it to dynamically adjust the learning rate. At each iteration, for the gradient g_i of the parameter θ_i, AdaGrad calculates the update using the following formula:

$$\Delta\theta_i = -\frac{\eta}{G_i + \varepsilon} g_i \tag{9}$$

where η is the learning rate and ε is a very small constant that ensures the denominator is non-zero-valued, and after each update, AdaGrad updates G_i to the cumulative sum of squared historical gradients:

$$G_i = G_{i-1} + g_i^2 \tag{10}$$

Since the learning rate of each parameter is dynamically adjusted, AdaGrad is suitable for a wide range of different problems and performs well especially when dealing with sparse data. However, the disadvantage of AdaGrad is that it will make the learning rate close to zero with the increase of time, leading to the early termination of the training process. In addition, for those parameters with large gradients, their accumulation will become larger and larger, resulting in a smaller and smaller learning rate, which in turn leads to a decrease in the optimization speed, which is also a limitation of the AdaGrad optimizer.

(2) Momentum gradient descent is a commonly used gradient descent algorithm, and the core idea is to add momentum to the gradient descent process. Momentum can be understood as the inertia of gradient descent, which serves to maintain the original direction and accelerate the gradient descent process. In each iteration, the Momentum optimizer updates the weights according to the direction of the current gradient and the last gradient, so that the weights are accumulated in the direction of the gradient and the gradients in different directions are smoothed. This makes the optimization algorithm more stable, avoiding oscillations and jitter, and also speeds up the convergence of the algorithm.

The formula for updating the parameters of the Momentum optimizer is as follows:

$$\begin{cases} v_t = -\beta\nabla_\theta J(\theta_t) + (1-\beta)v_{t-1} \\ \theta_{t+1} = \theta_t + v_t \end{cases} \tag{11}$$

where β represents the learning rate, θ_t denotes the parameter at the moment t, $J(\theta_t)$ represents the loss function at the moment t, and v_t represents the updating rate of the parameter at the moment t. v_{t-1} represents the update rate of the parameter at the previous moment.The advantage of the Momentum optimizer is that it can quickly converge to the global optimal solution and can avoid the problem of oscillations and jitter in the gradient descent process. In addition, the Momentum optimizer can help the algorithm to jump out of the local optimum solution and can better handle non-convex functions. Compared with other optimization algorithms, the Momentum optimizer is relatively simple to implement, and the performance and convergence speed of the algorithm can also be controlled by adjusting the parameters. However, the Momentum optimizer also has some drawbacks. First, due to the introduction of momentum, the Momentum optimizer takes a longer time to adapt to the new data distribution. Second, if the algorithm converges to a suboptimal solution, the Momentum optimizer will continue to oscillate around that suboptimal solution. Therefore, in practice, the parameters of the Momentum optimizer need to be carefully tuned so that the algorithm can better converge to the global optimal solution.

(3) RMSProp is also a commonly used gradient descent algorithm, its full name is Root Mean Square Propagation (RMSP), which is mainly used in the training of neural networks.The basic idea of the RMSProp optimizer is to take a weighted average of the gradient of each parameter and use it as the learning rate at the time of update. Specifically, RMSProp uses an exponentially weighted average to compute the root mean square of the gradient and divides it by the learning rate as the scaling factor when the parameter is updated.The benefit of the RMSProp optimizer is that it can adaptively adjust the learning rate, which can be different for different parameters, and it can effectively reduce the dependency between the neurons, which makes the gradient become more stable and controllable, avoiding the tendency of gradient vanishing and gradient explosion.The parameter update formula of RMSProp optimizer is as follows:

$$\begin{cases} s_t = \beta s_{t-1} + (1-\beta) g^2 \\ \Delta\theta = -\dfrac{\eta}{\sqrt{s_t + \varepsilon}} g \end{cases} \tag{12}$$

where β is the learning rate, ε is a very small constant that ensures the denominator is non-zero, s_t state variable is an exponentially weighted moving average over the squared term g^2, and θ is the parameter to be optimized.

(4) Adam is an optimization algorithm based on gradient descent, which is one of the optimization algorithms that have been widely used in the field of deep learning in recent years.The Adam algorithm combines the advantages of the AdaGrad and RMSProp optimizers, and adds the mechanism of bias correction, which is capable of adaptively adjusting the learning rate, and is able to adjust the learning rate dynamically during the training process.The Adam optimizer's The core idea is to update the model parameters based on the first-order moment and second-order moment estimation of the gradient. In the Adam optimizer, the first-order moment estimation is the exponentially weighted average of the gradient, and the second-order moment estimation is the exponentially weighted average of the gradient squares. Specifically, for each parameter gradient, the Adam algorithm computes their first-order moments and second-order moment estimates and updates the model parameters by dividing them by their respective exponentially weighted averages.The Adam optimizer parameter updating formula is as follows:

$$\begin{cases} m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \\ v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \\ \hat{m}_t = \dfrac{m_t}{1-\beta_1^t} \\ \hat{v}_t = \dfrac{v_t}{1-\beta_2^t} \\ \Delta\theta = -\dfrac{\eta}{\sqrt{\hat{v}_t + \varepsilon}} \hat{m}_t \end{cases} \tag{13}$$

where g_t is the gradient, η is the learning rate, β_1 and β_2 are the decay rates of the momentum and squared gradient, respectively, m_t and v_t are the estimation of the first-order and second-order momentum, respectively, and $\hat{m}_t$ and $\hat{v}_t$ are the bias corrections to the first and second order momentum, θ is the parameter to be optimized, and ε is a small constant preventing the denominator from being zero.

In updating the parameters according to the above equation (13), Adam's algorithm uses stochastic gradient descent with momentum, which makes the update direction inertial and thus moves more smoothly in the parameter space. In addition, the Adam algorithm also corrects the first-order moment estimates and second-order moment estimates through a bias correction mechanism, which solves the problem of inaccurate gradient estimation at the beginning of training. Overall, Adam's algorithm is an efficient optimization algorithm with adaptive learning rate, momentum and bias correction, which is able to dynamically adjust the learning rate during the training process so as to better adapt to different data distributions and task requirements, and is one of the most commonly used optimization algorithms in the field of deep learning.

III. Analysis of the effect of music creation and teaching based on artificial intelligence

III. A. Evaluation of Artificial Intelligence for Music Composition

III. A. 1) Evaluation results

In this thesis, 30 professional listeners (music students) will be selected, this part of the audience has the experience of playing musical instruments or senior music lovers and 25 ordinary listeners, these 25 ordinary listeners do not have access to any musical instruments, so that they score the different works, the professional audience's scoring weight for the ordinary audience of 1.5 times, according to this scheme to find the average score, that is, for a piece of music The average score will be the manual scoring score of a particular piece of music.

Listeners will score the music according to their feelings and preferences. In order to ensure the fairness and accuracy of the scoring, a method called “weighted average” is used to calculate the score of each piece of music, with the weight of professional listeners being 1.5 times that of ordinary listeners. This method takes into account the opinions of different groups of listeners and calculates the final score based on their weights. Each listener needs to be considered holistically, e.g. melody, rhythm, harmony, etc. The final score is the result of human scoring of the music.

An experiment was conducted to compare the difference between AI scoring and human scoring. In this thesis, 10 brand-new MIDI music were selected and the listener was informed of the characteristics of the MIDI music in advance so that factors such as emotion and timbre were excluded, and then the music was given to the AI and the listener for scoring respectively. Table 1 shows a comparison of the scoring results.

There is also variability in the scoring between different listeners. Professional listeners give different scores to different music, and are more inclined to complex music, while ordinary listeners are more enthusiastic about bright melodies and chords, and are not inclined to give scores with a large gap. At the same time, it can be seen that although there is a certain degree of variability between the AI scores and the listener scores, the overall variability is not large, and the AI scores are much closer to the scores of the professional listeners, with the difference between the artificial scores and the model scores being about 1.5 points. This also shows that the scoring of ordinary listeners is more subjective, and is easily affected by personal emotional factors, or different preferences and other factors. It can be seen that AI scoring and manual scoring can complement each other to provide more comprehensive and accurate results for music scoring.

Table 1: Score contrast

Music	style	Professional audience score	Average audience score	Final score	Model score
MIDI1	Gloom	93.542	90.394	92.248	93.248
MIDI2	Gloom	93.348	91.648	92.699	94.148
MIDI3	Warmth	88.399	90.294	89.148	87.599
MIDI4	Warmth	93.493	93.634	93.488	92.548
MIDI5	Brightness	85.618	93.348	88.796	87.378
MIDI6	Brightness	88.969	93.548	90.745	90.169
MIDI7	Enthusiasm	89.168	91.264	89.896	90.148
MIDI8	Enthusiasm	90.045	91.248	90.548	90.248
MIDI9	Peace	90.163	90.496	90.264	90.048
MIDI10	Peace	94.598	91.148	93.144	93.348

III. A. 2) Composition model training evaluation

As mentioned in the previous chapter, if too much training results in a loss function that is too small, then the music may not be lively enough or may sound unchanged, and this is not a good piece of music. Therefore, when training the model, it is necessary to find an appropriate stopping time that enables the model to generate the best-sounding music.

Specifically, the stopping time of training is controlled by adjusting the number of iterations of the model. When the loss function reaches a certain threshold, it is sufficient to stop training and generate music. How to define this threshold then becomes a problem. Since there are already two ways of evaluating a piece of MIDI music, it is possible to stop training at different iterations and generate different tunes according to the same rules, and find the highest rated iteration by rating the different tunes. For each iteration number different tunes are selected to take their average score, Table 2 shows the music scoring for different iteration numbers, when the iteration number is 60k, the human score and model score both reach the highest value, 79.249 and 80 respectively, and the number of generated tunes is 5.

Table 2: Different iterations of the music score

Iteration number	Artificial score	Model score	Number of music
10k	Ungraded	/	1
15k	Ungraded	64.518	2
20k	61.348	64.398	4
25k	63.288	66.548	5
30k	65.539	69.365	7
35k	73.569	72.548	7
40k	75.648	77.948	7
50k	76.445	75	8
60k	79.249	80	5
80k	73.248	73.445	6
100k	73.048	73.248	5

III. A. 3) Comparison of composing results of different models

The Keras deep learning framework was used for the AI AI composition and final generation of MIDI music. A certain amount of MIDI experimental material is used in the experimental process, and the neural network is used to allow the AI to train and generate the neural network model based on the provided music material, and finally generate the composing model.

In order to compare the effect of DNN model composing with traditional neural network model composing, this thesis uses the two models to amplify the music using the same musical phrases respectively, and scores the final generated music. The musical phrases were selected from 1 second to 32 seconds.

Table 3 shows a comparison of the models, with six short musical phrases of different lengths selected for amplification. It can be seen that the longer the original musical phrase, the higher the score of the music made by the two models when music amplification is performed, and it can also be seen that the performance of the model used in this thesis is better than the performance of the music made by using the Keras deep learning framework, and the general score is higher by more than 10 points, and the mean value of the score of the DNN model is 77.063 points.

Table 3: Model comparison

Music	Duration	Keras Frame score	DNN Frame score	Fractional difference
Music section 1	1s	63.248	76.288	13.04
Music section 2	3s	64.399	75.596	11.197
Music section 3	8s	65.128	77.248	12.12
Music section 4	18s	65.299	77.399	12.1
Music section 5	28s	64.948	77.048	12.1
Music section 6	32s	66.768	78.798	12.03

III. B. Evaluation of the effectiveness of music teaching

III. B. 1) Evaluation of teaching and learning

In this lesson, the teacher used two kinds of teaching evaluation: peer evaluation and teacher evaluation, but mainly teacher evaluation, which was designed by the author based on the content of the music teaching test, and summarized the classroom assessment as shown in Table 4.

There are 30 students in the class, of which 23 students scored more than 85 points, accounting for 76.67%, 7 students scored 60-85 points, accounting for 23.33%, and 0 students scored less than 60 points. it can be seen that the class attendance rate is high, and there is no student absenteeism, and the students' participation is very high, the AI-based music classroom teaching mode makes the teacher form an equal and interactive relationship with the students, the students dare to speak their minds, and the students are more willing to speak their minds in accordance with the teacher evaluation. The AI-based music classroom teaching mode makes teachers and students form an equal interactive relationship, and students dare to speak out their own ideas and express music according to their own favorite way, so the classroom atmosphere encourages teachers and students to learn, discuss problems and solve problems together.

Table 4: Evaluation results

ID	Music information extraction	Music data processing	Music language expression	Music exploration process	Music thinking diverges	Music knowledge migration	Total score
1	3	3	3	2	3	3	2
2	3	2	3	3	3	3	3
3	3	2	3	2	3	2	2
4	2	3	3	2	2	2	3
5	3	3	2	3	3	2	3
6	3	3	3	3	2	3	3
7	2	3	3	3	3	3	3
8	3	3	2	2	3	3	3
9	3	3	3	3	3	2	3
10	2	3	2	3	3	3	3
11	3	3	3	2	3	3	3
12	3	3	3	2	2	3	3
13	3	3	3	3	2	3	3
13	3	3	3	2	3	3	2
15	3	3	3	3	2	3	3
16	2	2	3	3	3	3	2
17	2	2	2	3	3	2	3
18	2	3	3	3	3	3	3
19	3	3	2	3	3	2	3
20	2	3	3	3	2	3	2
21	2	3	3	3	2	2	3
22	2	3	3	2	3	3	3
23	2	3	3	3	3	3	3
23	3	3	3	3	3	3	3
25	3	2	3	2	3	3	3
26	3	3	2	3	3	2	3
27	3	3	3	3	3	2	2
28	3	2	3	3	2	3	3
29	3	3	2	2	2	3	2
30	3	3	3	3	2	2	3

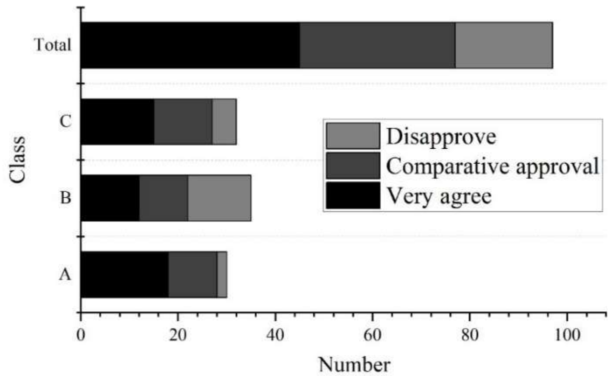
Note: The evaluation criterion is:
1=(less than 60 points)
2= (60——85 points)
3= (85 points or above)

III. B. 2) Acceptance of AI music instruction

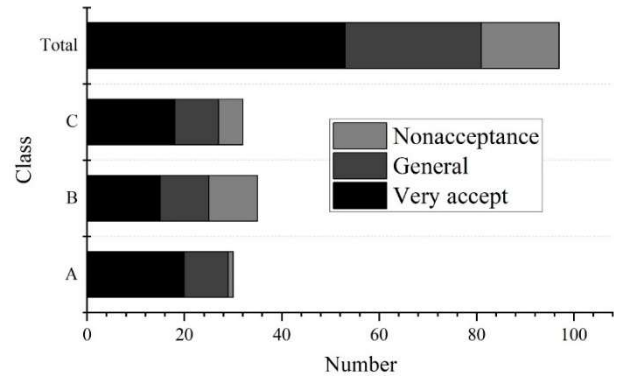
The 30 students who have gone through AI music teaching are divided into class A. Now 35 students in class B (traditional teaching) and 32 students in class C (mixed teaching) are added. After a semester of music class, it helps to design AI music classroom teaching that meets the students' learning situation according to the psychological characteristics of music learning of students at this age.

Figure 3 shows the students' acceptance of AI music classroom, Figure (a) is the approval, Figure (b) is the model acceptance, it can be seen that most of the students still agree with the AI classroom as a teaching model. 46.392% of the students strongly agree with the flipped classroom teaching model, 32.99% are more in favor of the flipped classroom, and 20.618% of the students do not agree with the flipped classroom. The highest number of students in class A agreed with it, 60% of the class strongly agreed with it, 33.333% of the class somewhat agreed with it, and only 2 students in the class disagreed with the implementation of AI classroom teaching.

54.64% of the students were very receptive to the AI modeling of music composition, and 28.866% of the students felt that the AI modeling of learning was average.



(a)Approbation



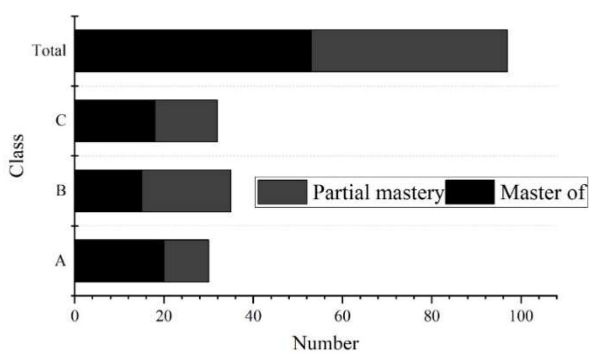
(b)Model acceptance

Figure 3: Student acceptance of ai music hall

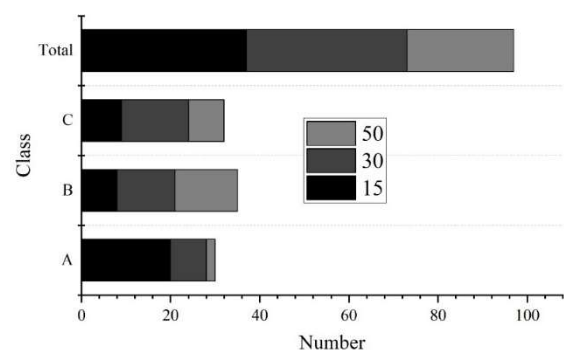
III. B. 3) Degree of mastery of teaching content

Figure 4 shows the students' mastery of the AI teaching content, Figure (a) shows the mastery of the learning content, Figure (b) shows the time spent on mastering the content, after different types of music classes, 54.692% of the students can completely master the learning content, and 45.308% of the students can master a part of the learning content. Among them, Class A has the largest number of students who completely mastered the learning content, with 20 students. After the teacher's investigation, it was found that the music committee of this class spontaneously organized the students of this class to collectively carry out the second study before the AI model class after the class, so the students' mastery was higher than that of the other two teaching classes.

24.742% of the students could finish the learning content in 50 minutes, 37.113% of the students spent 30 minutes on learning, and 38.163% of the students could finish the learning in a shorter period of time. The number of students in class A who could finish the learning in 15 minutes was 20, which was significantly higher than that of the other classes, which indicated that there was a variability in the learning ability of each class.



(a)Learning content



(b)The time spent in mastering the content

Figure 4: Students' knowledge of the content of ai teaching

IV. Conclusion

In this study, the application of artificial intelligence in music composition and teaching showed significant and innovative effects. For the evaluation of musical compositions, the difference between the AI ratings and the ratings of professional listeners was only 1.5 points, indicating that the AI was able to more accurately mimic the perceptions of professional listeners in evaluating musical compositions. In addition, the model training shows that with the increase in the number of training times, the score of the music work gradually improves, and when the number of iterations is 60k, the artificial score reaches 79.249 points, and the model score is 80 points, which indicates that the AI model is able to generate a higher-quality work in the process of music creation.

In music teaching, the personalized teaching model based on artificial intelligence effectively improves students' learning interest and efficiency. Experimental data show that students who have been taught by AI are able to master the music content in a shorter period of time, in which the proportion of students in class A who finish learning within 15 minutes is significantly higher than that of other classes. In addition, students' acceptance of the AI music classroom is high, with more than 54% of students indicating that they strongly agree with this teaching mode, indicating that the application of AI in music education has a broad prospect.

Funding

This research was supported by the Second Batch of 14th Five-Year Plan Undergraduate Teaching Reform Projects in Zhejiang Province in 2024: Exploration and Practice of Digital Teaching System for Music Design and Production Major Empowered by Artificial Intelligence.

References

[1] Gioti, A. M. (2020). From artificial to extended intelligence in music composition. Organised Sound, 25(1), 25-32.
[2] Xin, C. (2024, February). Research on Music Creation System Based on Computer Artificial Intelligence Technology. In 2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA) (pp. 941-945). IEEE.
[3] Li, N., & Cai, J. (2023). The Aesthetic Analysis of Music Generation Algorithms Based on Artificial Intelligence Technologies. Humanities and Social Sciences, 11(6), 223-230.
[4] Ventayen, R. J. M. (2024). Music Creation by Artificial Intelligence and Listener Reactions in a Blind Test. Available at SSRN 5023874.
[5] Morreale, F. (2021). Where does the buck stop? Ethical and political issues with AI in music creation. Transactions of the International Society for Music Information Retrieval, 4(1).
[6] Carnovalini, F., & Rodà, A. (2020). Computational creativity and music generation systems: An introduction to the state of the art. Frontiers in Artificial Intelligence, 3, 14.
[7] Pfeiffer, A., & Wernbacher, T. (2025). ARTIFICIAL INTELLIGENCE AND MUSIC: EVOLUTION AND INTEGRATION IN UPPER SECONDARY MUSIC EDUCATION IN AUSTRIA. In INTED2025 Proceedings (pp. 262-267). IATED.
[8] Pereverzeva, M. V. (2021). The prospects of applying artificial intelligence in musical composition. Russian Musicology, (1), 8-16.
[9] Nugroho, Y. Y. T., & Manggala, P. P. M. D. (2024, December). The Use of AI in Creating Music Compositions: A Case Study on Suno Application. In 7th Celt International Conference (CIC 2024) (pp. 177-189). Atlantis Press.
[10] Tigre Moura, F., & Maw, C. (2021). Artificial intelligence became Beethoven: how do listeners and music professionals perceive artificially composed music?. Journal of Consumer Marketing, 38(2), 137-146.
[11] Alaeddine, M., & Tannoury, A. (2021). Artificial intelligence in music composition. In Artificial Intelligence Applications and Innovations: 17th IFIP WG 12.5 International Conference, AIAI 2021, Hersonissos, Crete, Greece, June 25–27, 2021, Proceedings 17 (pp. 387-397). Springer International Publishing.
[12] Louie, R., Coenen, A., Huang, C. Z., Terry, M., & Cai, C. J. (2020, April). Novice-AI music co-creation via AI-steering tools for deep generative models. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-13).
[13] Holland, S. (2013). Artificial intelligence in music education: A critical review. Readings in music and artificial intelligence, 239-274.
[14] Boosa, S. (2019). AI and Creativity: How Writing, Music, and Art Are Increasingly Revolutionized by Machine Learning. EPH-International Journal of Science And Engineering, 5(4), 37-47.
[15] Mycka, J., & Mańdziuk, J. (2024). Artificial intelligence in music: recent trends and challenges. Neural Computing and Applications, 1-39.
[16] Cipta, F., Sukmayadi, Y., Milyartini, R., & Hardini, T. I. (2024). Optimizing AI-Powered Music Creation Social Media to Amplify Learning Content. Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran, 10(3), 881-892.
[17] Kumar, L., Goyal, P., & Kumar, R. (2020). Creativity in machines: music composition using artificial intelligence. Asian Journal For Convergence In Technology (AJCT) ISSN-2350-1146, 6(2), 36-40.