Environment and Social Psychology (2024) Volume 9 Issue 11
doi: 10.59429/esp.v9i11.3081

Research Article

Analyzing AI composition techniques and their influence on human musical aesthetics using bi-GRU and self-attention models

Qi Gao[1,*], Jinting Cai[1], Fuxin Wang[2], Junsong Chang[3], He Huang[4]

[1] *Graduate School of Global Culture Convergence,Kangwon National University,Chuncheon City,Gangwon Province,24341,South Korea*

[2] *Jiangxi Agricultural University, Nanchang, Jiangxi, 330045, China*

[3] *College of Art, Kyung Hee University, Seoul Special City, Seoul Special City, 02447, South Korea*

[4] *College of Music, Qinghai Normal University, Xining, Qinghai Province, 810000， China*

***Corresponding author:** Qi Gao, ArtGaoqi@163.com

ABSTRACT

With the continuous development of artificial intelligence technology, this study aims to explore the application of artificial intelligence in human aesthetics. By processing the signal in sub frames and using short-time Fourier transform to analyze the position information of beat points, the start and key musical features of notes can be accurately detected. Based on the extracted music features, a Bi GRU network and self-attention mechanism automatic composition model are established to process important information between longer sequence predictions and prominent notes, and to evaluate the accuracy and vividness of AI composed music works. The results showed that the model achieved an accuracy of 94.28% in processing melody and rhythm data. Excellent performance in terms of music fluency and coordination, with high scores in human music aesthetics indicators, reaching a pitch score of 92, and classical style scores of 90 and 92 in melody and integrity. Artificial intelligence has to some extent influenced and shaped human music aesthetics, providing important evidence for understanding its impact on music creation.

Keywords: beat point position; music characteristics; automatic composition; music fluency; musical composition

1. Introduction

AI composition is also called AI music composition, and with the development and application of AI technology, the degree of automation of music generation systems can be categorized into two types[1]. One is the more traditional one, which uses machine learning methods as a music generation model, and its generation system has a lower degree of automation[2]. Such as computer-assisted composing or computer-assisted composition, the overall creative activity of music is more dependent on the composer[3]. The second is more popular nowadays, which uses deep learning techniques as generative models, and its generative system is more automated[4]. Systems such as complete generation, partial generation, or human-computer organic interaction in human-computer collaboration are capable of automatically generating complete or

ARTICLE INFO
Received: 5 September 2024 | Accepted: 15 November 2024 | Available online: 25 November 2024

CITATION
Gao Q, Cai JT, Wang FX et al. Analyzing AI composition techniques and their influence on human musical aesthetics using bi-GRU and self-attention models. *Environment and Social Psychology* 2024; 9(11): 3081. doi:10.59429/esp.v9i11.3081

COPYRIGHT
Copyright © 2024 by author(s). *Environment and Social Psychology* is published by Arts and Science Press Pte. Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), permitting distribution and reproduction in any medium, provided the original work is cited.

more complete music compositions. Based on the higher degree of automation of the music generation system, according to the different degree of human-computer interaction, there are two specific methods of division of labor in human-computer collaboration, one is the systematic autonomous generation of the machine, and the other is the compositional collaboration[5]. Therefore, in the present time, the collaboration between human artists and AI to generate art works will be a creative process full of infinite possibilities, and at the same time, the artists can utilize their own creativity to improve the artistic value of the generated works[6].

In recent years, with the continuous progress of AI technology, AI composition has become a special way of composing, and many scholars have begun to carry out a lot of research on it. Gioti argues in his research that AI technology can generate many different styles of music in a short time, which not only avoids the problem of bottlenecks that are often encountered by human composers, but also is more versatile in terms of composing styles, and brings different experiences to the listeners by combining many different elements together, and promotes the progress and diversification of human compositions through the advantages of AI composition[7]. AI technology can generate many different styles of music in a short time, not only avoiding the bottleneck problem often encountered by human composers, but also being more versatile in terms of composition styles, combining many different elements together, bringing listeners different experiences, and facilitating the progress and diversification of human compositions through the advantages of AI compositions. Morreale analyzed a large amount of AI compositions, and found that the other advantage of AI compositions is the advantage of customization, that is, compositions are composed according to the needs of the listener. demands, and produce music that meets the listener's aesthetic in a short time, promoting the listener's personalized experience of music[8]. Pachet et al. adopted the network behavioral characteristics of biological nerves to mimic the relevant mathematical model, and completed the programming based on constructing a neural network, which inputs the music-related feature data into the neural network, and the internal mathematical model will Analyze the features of the music data[9]. Through the classification of features, the original melody, rhythm, and intensity state of the music is strengthened to give the listener a more distinctive experience.

Dai believes that the main advantage of AI composition is that it can learn the global characteristics of musical works and create more similar or even more unique works based on the training of a large number of samples, so that the listener can feel a variety of musical elements of human compositions when appreciating AI compositions[10]. At the same time can also feel the fusion of fresh elements, AI composition maintains both the tradition of human composition and the innovative type of AI composition. Deruty believes that the essence of AI composition is to extract and recognize the features of human composition through AI technology, and at the same time produce a melody of an innovative style. Taking note selection as an example, the notes are selected in a conversion table that meets the relevant standards, and combining the selected multiple notes together can form a creative thinking similar to human composition, with the disadvantage of the lack of a model of the work's melodic structure[11]. Briot points out in his study that AI technology is constantly researching and innovating in terms of imitating human emotions. However, it still can't have human emotions, and the creation of music needs the injection of emotions, human compositions can perfectly show the sadness in the music[12]. AI compositions, on the other hand, can't let the listener feel the existence of this emotion, and it is difficult to cause the emotional resonance of the listener's deepest heart. Huertas-Abril education is the key to overcoming differences, therefore social, cultural, and educational systems should be based on equality, educating more tolerant people, and ensuring respect for human rights[13]. Sircar proposed the latest progress in using machine learning and artificial intelligence to solve problems in the oil and gas industry. Helps eliminate risk factors and maintenance costs[14].

AI composition is not simply about generating notes and melodies, but more importantly about creating works that touch the heart and are in line with human musical aesthetics. Although there is a certain degree of certainty in related research, the performance of automatic composition is poor, and there is relatively little research on the impact of music style on human aesthetics. This article introduces the signal processing technology for analyzing AI music creation, and deeply analyzes the technical principles of AI composition, including music signal framing, short-time Fourier transform of audio signals, and complex domain spectral difference algorithm to generate starting point detection sequences, determine the extraction and representation of music features., Constructing a music evaluation model based on Bi GRU network and self attention mechanism, including the structure of automatic composition model, automatic composition training process, and composition evaluation process; Introduce self attention mechanism to flexibly adjust the dependency relationship between music features by assigning different weights. We have empirically verified the connection between AI composition and human music aesthetics, quantitatively analyzed which music features are more popular among humans, and explored the advantages of AI composition in different styles and emotional expressions. Provide theoretical support for new developments in the field of music creation. Revealing the complex relationship between AI composition and human music aesthetics, as well as its impact on human music aesthetics. At the same time, identify the limitations of the research and future research directions.

2. AI creation of music audio signal processing technology

2.1. Music signal framing

In this paper, mainstream music platforms such as NetEase Cloud Music and QQ Music, as well as streaming services such as Spotify and Apple Music are used to collect music works with AI creation or specialized AI music boards. The adopted music signals are MP3 format files with 16 KHz sampling rate and 128kbps encoding rate partially downloaded from the Internet, and some of the AI-created music is synthesized from recording devices and software, which may contain noise components. The proposed method of noise filtering based on cochlear impulse filter bank, after removing the noise from the audio, the signal can be further processed in a finer way by decomposing it into smaller analyzing units, which is the signal subframing to be introduced next.

Signal framing process is shown in **Figure 1**, from the time domain waveform of the music signal can be seen, the music signal belongs to the nonlinear signal, direct FFT processing of the signal will have a large distortion, which will have a greater impact on the timbre analysis. While the signal can be processed after framing can minimize this effect, it is generally considered that an audio signal is smooth on a 20-30ms segment[15].

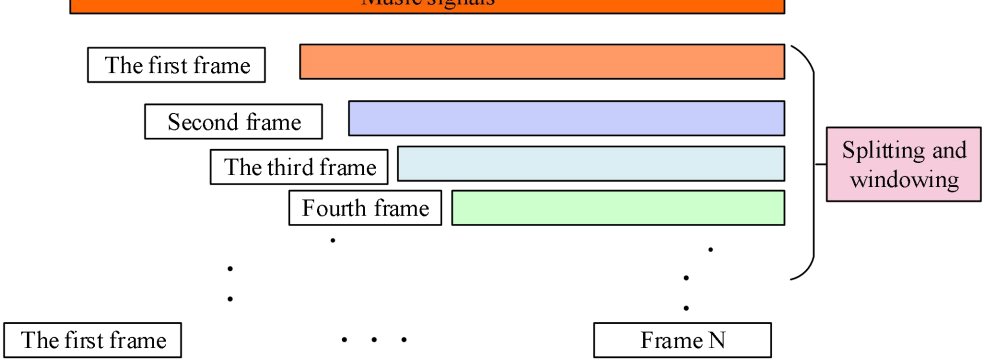


Figure 1. Signal framing process.

2.2. Short-time fourier transform of audio signals

After framesplitting, an in-depth spectral analysis of each frame is required to extract key features of the music. This requires the use of the short-time Fourier transform, a core technique in audio signal processing that is capable of transforming audio signals in the time domain into the frequency domain for analysis[16].

For musical signals, which are often accompanied by rapid changes in amplitude and frequency over time, capturing such changes is extremely important for the prediction of musical characteristics[17]. The short-time Fourier transform of a continuous-time signal is defined as follows:

$$STFT_x(t,\Omega)=\int_{-\infty}^{\infty}x(\tau)w(\tau-t)e^{-j\Omega t}d\tau \tag{1}$$

where $x(t)$ is the continuous time domain signal and $w(t)$ is the continuous window function. Similarly, for digital audio signals, the music signal is analyzed using a discrete form of the short-time Fourier transform, defined in the following form:

$$STFT_x(n,k)=\sum_{m=0}^{N-1}x(n+m)w(m)e^{-j\frac{2\pi}{N}mk} \tag{2}$$

The length of the window function N and the resolution of the short-time Fourier transform in the frequency domain is then determined. This resolution reflects the smallest frequency interval over which the transform can discriminate, and is critical for accurate extraction of musical features[18]. Combining the time-domain signal $x(n)$ and the window function $w(n)$ enables high-resolution spectral analysis of the audio signal to support subsequent AI composition and music aesthetic analysis. The expression form is:

$$\Delta f=\frac{f_s}{N} \tag{3}$$

Where f_s is the sampling frequency when converting from analog to digital signals, which is mostly fixed. However, a jump factor h is introduced in the short-time Fourier transform to point out the number of sample points each time the window function moves forward, corrected as follows:

$$S_k(m)=\sum_{n=0}^{N-1}x(n)w(mh-n)e^{-\frac{j2\pi nk}{N}} \tag{4}$$

Here, $S_k(m)$ is the short time Fourier transform of the m nd frame of sampled data and k is the frequency variable. In this paper, we use this equation to do a time-frequency analysis of the raw audio data and carefully and discretionarily choose the jump factor h.

2.3. Complex domain spectral difference algorithm to generate starting point detection sequence

In music signal analysis, accurate detection of note onsets is crucial for subsequent music feature extraction and rhythm analysis. The complex-domain spectral difference algorithm is an effective method for onset detection, and the onset detection function plays a key role in music information acquisition, connecting the original signal with the high-level music feature sequences, which is a kind of intermediary signal. Figure 2 shows the complex domain spectral difference algorithm to generate the onset detection, the algorithm over the calculation of the difference between two adjacent frames of the spectrum to identify the mutation points in the audio signal, which often correspond to the starting position of the note. The shape of the envelope of the ODF closely resembles that of the original signal, and each data point in the ODF sequence represents the likelihood of a musical event. Locating the time value and position of each note more precisely provides more accurate timestamp information for subsequent AI compositions.

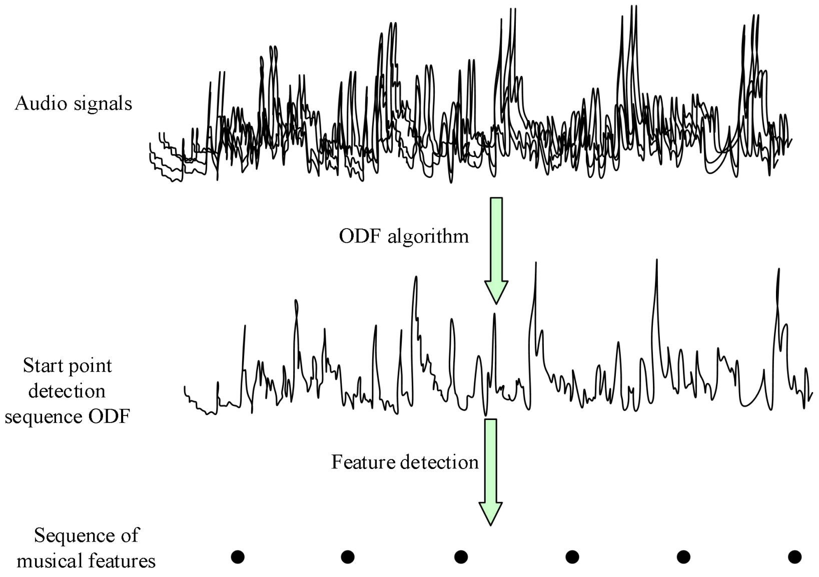

Audio signals
ODF algorithm
Start point detection sequence ODF
Feature detection
Sequence of musical features

Figure 2. Complex domain spectral difference algorithm to generate onset detection.

Accurate time-frequency analysis of the audio signal is a crucial step in the process of creating music in AI[19]. Next, the short-time Fourier transform is done for each frame, and the window function chosen here is the Hanning window, which can be selected from the actual program. The short-time Fourier transform of frame m of audio signal $x(n)$ is rewritten as follows:

$$S_k(m) = \sum_{n=0}^{N-1} x(n)w(mh-n)e^{-j\frac{2\pi}{N}nk} \tag{5}$$

Where $k = 0,1,\ldots,N-1$ represents each frequency component after Fourier transform, $h = 512$ is the jump factor, i.e., the number of sampling points that the window function moves forward on the original signal at one time. $w(n)$ is the Hanning window to improve the accuracy of spectral analysis. The Hanning window has better performance in both time and frequency domains and can effectively reduce the spectral leakage[20]. Its expression is as follows:

$$w(n) = 0.5 - 0.5\cos\left(\frac{2\pi n}{N}\right), 0 \le n \le N-1 \tag{6}$$

The difference between the short-time Fourier transform spectrum of frame m and the predicted spectrum of frame m yields a complex-domain spectral difference vector, whose modulus squared constitutes the onset detection function, reflecting the degree of difference between the predicted information and the actual observed information:

$$CSD(m) = \sum_{k=0}^{N-1} \left| S_k(m) - \hat{S}_k(m) \right|^2 \tag{7}$$

The above equation describes the degree of similarity between the predicted information and the observed information, the larger the value obtained the less similarity, the more likely the musical event will occur[21]. Since the discrete Fourier transform spectrum of the real signal satisfies the characteristics of even symmetry of amplitude-frequency response and odd symmetry of phase-frequency response, only the spectral information at $k = 0,1,\ldots,N/2-1$ is taken here to compute the value of the onset detection function $\Gamma(m)$ for the audio signal in frame m, i.e.:

$$\Gamma(m)=\sum_{k=0}^{N/2-1}\left|S_k(m)-\hat{S}_k(m)\right|^2 \tag{8}$$

In order to adapt to changes in the speed and phase of the music, here at the starting point detection function $\Gamma(m)$ uploads a rectangular window of length $B=512$ for observation, and each time the know-shaped window jumps forward to a length of $k=0,1,\ldots,N/2-1$, that is, there:

$$\Gamma_i(m)=\Gamma(m),\quad m=1+(i-1)B_f,\ldots,B_f+(i-1)B_h \tag{9}$$

Where, $B_f=512$ length, $\Gamma_i(m)$ is a start point detection frame, music rhythm information body acquisition are processed with the detection frame.

3. Bi-GRU network and self-attention mechanism for music assessment modeling

In order to evaluate the quality of the musical compositions composed by AI, an evaluation model based on Bi-GRU, called Bi-Directional Gated Recurrent Unit Network and Self-Attention Mechanism, is constructed. The model is able to comprehensively capture the temporal features and internal structures in music sequences, providing objective evaluation criteria for AI compositions. The gated recurrent unit network is able to solve the gradient vanishing and gradient explosion problems of recurrent neural networks as well as deal with the long-term dependency problem in sequence prediction. In addition, music is a kind of expressive art with certain melody and rhythm that changes with the ebb and flow of time, and it usually needs to perceive the temporal information of the music and relate it to the contextual semantics of the music to be able to accurately classify the musical works. Therefore, this paper chooses the structure of bidirectional recurrent network to deeply portray the contextual timing information of music, and considers not only the influence of historical note information but also the influence of future note information when generating the current note. At the same time, by adding the self-attention mechanism to automatically assign different probability weights to the note features in the context, the internal dependency features of the note sequences can be deeply explored, and the note sequences can be better expressed, so as to improve the accuracy of the assessment of the note sequences.

3.1. Structure of the autocomposition model

Based on the extracted musical features, an automatic composition model was further constructed. By training the model to generate musical compositions with specific styles and emotions, the potential and innovation of AI in music composition can be explored. At the same time, this automatic composition model will also provide a large number of samples of musical compositions for subsequent evaluation and analysis.

Figure 3 Structure of the automatic composing model, it can be seen that the whole composing model can be divided into three pieces from the big module: input layer, hidden layer and output layer. x represents the input layer data, O represents the output layer data, and h represents the hidden layer data. The input and output modes are chosen to be synchronized many-to-many modes, i.e., the input note sequences and the output note sequences are both more than one note and of equal length. It consists of two main parts of the work:

(1) MIDI main melody note sequences are acquired as note feature vectors with contextual information.

(2) Bi-GRU stores the timing dependency information of the input sequences through an internal bi-directional structure, and learns to compute the impact on the current moment state from both future and historical directions respectively. Then the output of the Bi-GRU network is used as the input of the self-attention mechanism layer to obtain the corresponding probability distribution of the attention weights. The

representation of the note sequence vectors after adding the self-attention mechanism uses the sofmax function to map the output of the hidden layer to the probability distribution of the note numbers, and for each note vector in the output note sequence, the note number with the highest probability is selected as the final predicted note.

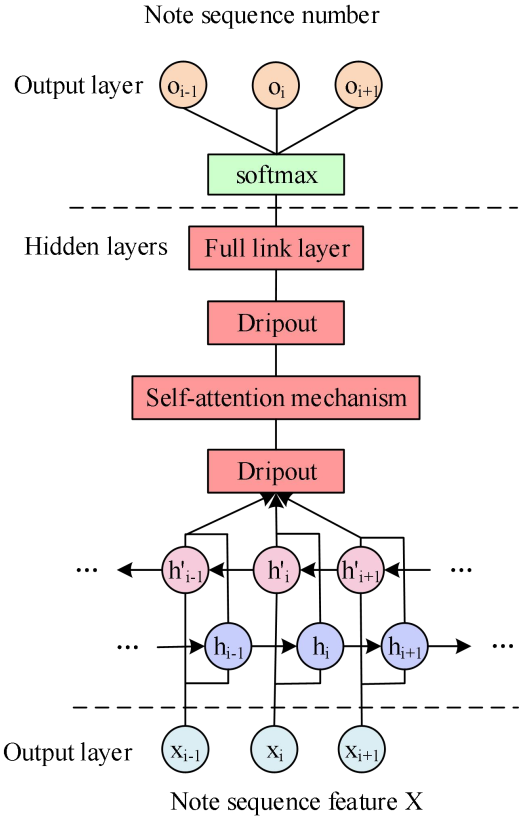


Figure 3. Structure of the autocomposition model.

3.2. Automatic composition training process

The data used by the automatic composition model, i.e., the main melody note sequences, comes from MIDI main melody files, and in this paper, we use the main melody extraction method based on contour line algorithm and average pitch to extract all the MIDI main melody files. The main melody note sequences are represented by note feature vectors with contextual information, and this paper adopts the method of generating note feature vectors based on contextual semantic coding to obtain the note feature vectors. The main melody note sequence data is divided into a training set, a validation set, and a test set in the ratio of 7:2:1. During the training process of the automatic composition model, the training set data is used for training to learn the composition knowledge, while the validation set is periodically used to verify the effect of the current model training during the training process. That is, the current validation set note prediction accuracy effect, the test set is the data set used to predict notes in the composition process of automatic composition. The entire model uses the synchronized many-to-many input-output mode of the recurrent neural network, and the input of the autocomposition model is a note sequence of length " , and the output is also a note sequence of length " . Therefore, the main melody note sequence needs to be sliced, and the main melody note sequence is defined as $Notes = [note_1, note_2, \cdots note_t]$, then the matrix of the sliced note sequence X is as follows:

$$X = \begin{bmatrix} note_1 & note_2 & \cdots & note_n \\ note_{n+1} & note_{n+2} & \cdots & note_{2^* n} \\ \vdots & \vdots & \vdots & \vdots \\ note_{i^* n+1} & note_{i^* n+2} & \cdots & note_{(i+1)^* n} \end{bmatrix} \tag{10}$$

Define the desired output note matrix Y as follows:

$$Y = \begin{bmatrix} note_{n+1} & note_{n+2} & \cdots & note_{2^* n} \\ note_{2^* n+1} & note_{2^* n+2} & \cdots & note_{3^* n} \\ \vdots & \vdots & \vdots & \vdots \\ note_{(i+1)^* n+1} & note_{(i+1)^* n+2} & \cdots & note_{(i+2)^* n} \end{bmatrix} \tag{11}$$

The matrix, i.e., the sliced note sequence, is also the input data for the autocomposition model. The latter note sequence of the sliced current input note sequence is used as the desired output, i.e:

Current input note sequence $X_i = \left[\ note_{i^* n+1},\ note_{i^{**}+2}, \cdots,\ note_{(i+1)^* n}\ \right]$, desired output note sequence $Y_i = \left[\ note_{(i+1)^{n+1}},\ note_{(i+1)^{n+2}}, \cdots,\ note_{(i+2)^m}\ \right]$. The rules of the training process are as follows:

(1) The autocomposition model receives a fixed-length phrase note sequence, and the output predicts the next phrase note sequence.

(2) The desired output note sequences in the training set are compared with the predicted output, and the error between the two is calculated using the crossover function.

(3) Update the learning parameters in the auto-composition model through the back-propagation algorithm, and after several rounds of iterative training and learning until the model converges, a good note evaluation model is obtained.

3.3. Composition assessment process

Figure 4 shows the music evaluation of Bi-GRU network and self-attention mechanism, in the composition evaluation process, the AI-generated music piece is input into the evaluation model. The model will output a comprehensive score reflecting the overall performance of the work in terms of melody, rhythm, harmony, etc. This score can be used as an important basis for optimizing the AI music composition algorithm. The automatic composition model will eventually get a converged good note evaluation model after continuous training and learning. The note evaluation model also requires the input of a fixed-length musical note sequence, the first input note sequence is randomly selected from the test set, and then the note evaluation model outputs the predicted next fixed-length musical note sequence, and then the output musical note prediction sequence is used as the input to make the next prediction again. This iteration is repeated until the pre-set length of the generated phrase is generated, resulting in an automatically composed piece.

In the test set, each time a segment note sequence of fixed length is selected as input, a piece of music is generated by the note evaluation model, so that multiple pieces of music can be generated. The input dataset for the note evaluation model to generate a piece of music in the automatic composition model can be represented as follows:

$$X = \begin{bmatrix} note_1 & note_2 & \cdots & note_n \\ note_{pre_1} & note_{pre_2} & \cdots & note_{pre_n} \\ \vdots & \vdots & \vdots & \vdots \\ note_{pre_{i*n+1}} & note_{pre_{i*n+2}} & \cdots & note_{pre_{(i+1)*n}} \end{bmatrix} \tag{12}$$

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

$$Y = \begin{bmatrix} note_{pre_1} & note_{pre_2} & \cdots & note_{pre_n} \\ note_{pre_{n+1}} & note_{pre_{n+2}} & \cdots & note_{pre_{2+n}} \\ \vdots & \vdots & \vdots & \vdots \\ note_{pre_{(i+1)^{n+1}}} & note_{pre_{(i+1)^{n+2}}} & \cdots & note_{pre_{(i+2)^n}} \end{bmatrix} \tag{13}$$

The first row of data $X_1 = [note_1, note_2, \cdots, note_n]$ in X represents the sequence of musical notes randomly selected from the test set each time a musical composition is generated. The first line in the corresponding r is the generated musical note sequence $Y_1 = [note_{pre_1}, note_{pre_2}, ...note_{pre_n}]$, and then Y_1 is assigned to X_2 as the input for the next prediction, and the generation of musical note sequences is carried out iteratively in sequence, and finally the musical note sequences generated in r are spliced together by line to be the complete note sequences of the generated musical composition.

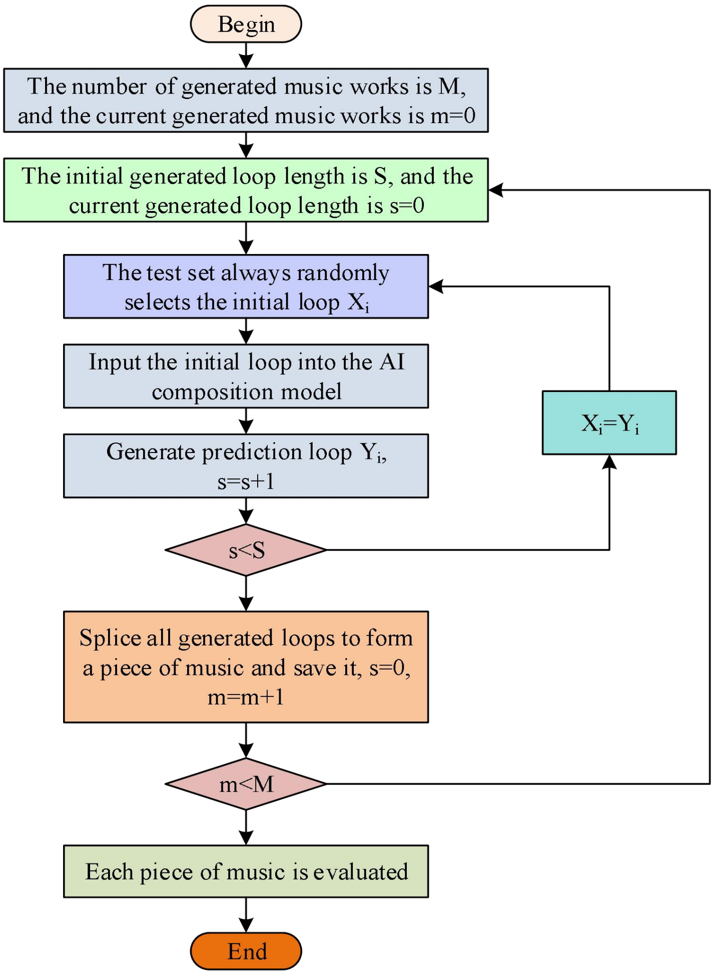

Begin
The number of generated music works is M, and the current generated music works is m=0
The initial generated loop length is S, and the current generated loop length is s=0
The test set always randomly selects the initial loop X_i
Input the initial loop into the AI composition model
Generate prediction loop Y_i, s=s+1
X_i=Y_i
s<S
Splice all generated loops to form a piece of music and save it, s=0, m=m+1
m<M
Each piece of music is evaluated
End

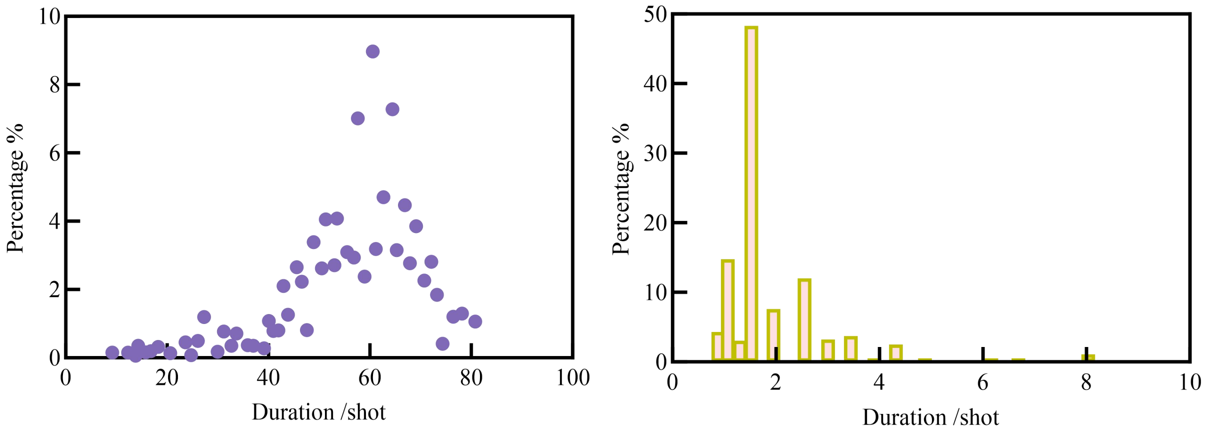
Figure 4. Bi-GRU network and self-attention mechanism for music evaluation.

4. The connection between AI compositions and human musical aesthetics

In order to achieve better test results, this paper obtains a total of 845 MIDI music files of the same style from the Internet, and obtains 510 qualified MIDI main melody music files after main melody extraction. All the MIDI music files are in 4/4 beat, the beat rate is 120 beats per minute, and a total of 33 kinds of notes are found. For the processing of melodic data, 20 lyric syllables were used as a fixed length of a single lyric to extract its corresponding melody. The melodic information was represented as a triad of pitch, duration, and rest time. If the length of the MIDI music was long, it was split, and the parts that were not long enough were discarded, yielding 11,149 lyrics with a length of 20 and the melody to its data. The data was split at a ratio of 0.8:0.1:0.1 to create training, validation and test sets.

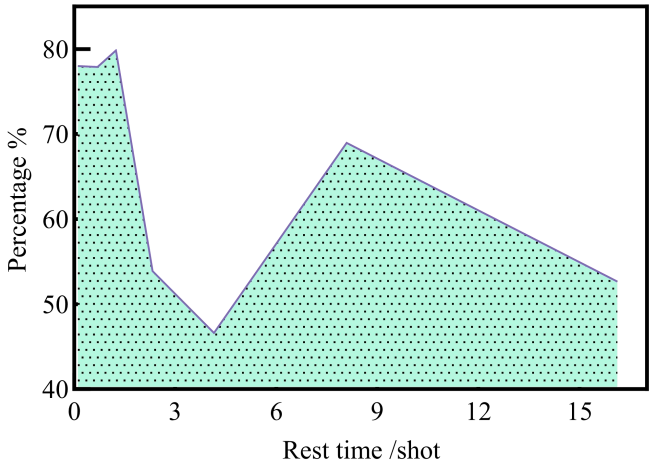
4.1. Distribution of automatically composed music attributes.

Figure 5 shows the distribution of the three attributes of the music modeled in this paper, and Figure 5(a) shows the pitch, with a theoretical value range of 21,108. In the model-generated music, the range of values in which pitch occurs more is concentrated between about 55,80. This finding is consistent with what happens in real music and verifies that the proposed model is able to simulate the pitch distribution of notes in real music well. It is also consistent with the general perception that most of the notes are in the mid-range and there are relatively few notes that are too high or too low. Figure 5(b) shows the durations, and the music generated by the model also shows a major concentration of less than two beats, with the percentage between 0-2 beats ranging from 4.3% to 48.3%, which indicates that the model is able to capture the note timings well. There are also a small number of notes whose durations are extended to vary from 4 to 8 beats, which usually appear at specific locations in the music, such as the end of a sentence or a place that needs to be emphasized, thus enhancing the expressive power of the music. Figure 5(c) shows the rest duration, and the percentage of model-generated music with a rest duration of 0 is close to 80%. This means that most of the notes are sung continuously, and rest time only occurs at the end of the phrase, at special pauses, or where a breath change is needed. This feature also matches the singing habits in real music, further demonstrating the accuracy of the model in simulating real music. To simulate end-of-sentence pauses or special pauses in real music, thus enhancing the rhythmic and expressive qualities of the generated music.



(a) Pitch

(b) Duration



(c) Body stop time

Figure 5. Distribution of three attributes of music modeled in this paper.

4.2. AI composition music quality accuracy

In order to verify the music matching and performance generated by the music generation model in this paper, the study utilizes the LakhDIMI dataset for simulation experiments. The data was preprocessed before the simulation experiments started and only the music with beat number 4/4 was retained. The learning rate of the model was set to 0.002 and the number of iterative training was 200. At the same time, the support vector machine SVM, bidirectional LSTM network, convolutional neural network CNN, multilayer perceptron MLP were compared with the generated model. The accuracy of AI composition music melody and rhythm processing is shown in **Figure 6**. The accuracy of melody and rhythm data processing of this paper's generative model is 94.28%, and the accuracy of melody and rhythm processing of the SVM is 84.96%. The SVM is used as a classical machine learning algorithm, has some generalization ability in dealing with classification problems. The melodic and rhythmic processing accuracy of bi-directional LSTM network is 83.01%, and bi-directional LSTM network, despite being able to deal with long-term dependencies in sequence data, did not manage to outperform the generative model in this experiment. The melodic and rhythmic processing accuracies of the convolutional neural network CNN and the multilayer perceptron MLP are 80.03% and 78.32%, respectively, for the CNN is better at processing data with spatial structure, while music data is more of a time series feature, while the MLP is unable to accurately identify the results of subtle changes in melodies and rhythms, which proves the generative model has an excellent ability to capture the complex melodies and rhythmic patterns with excellent capabilities.

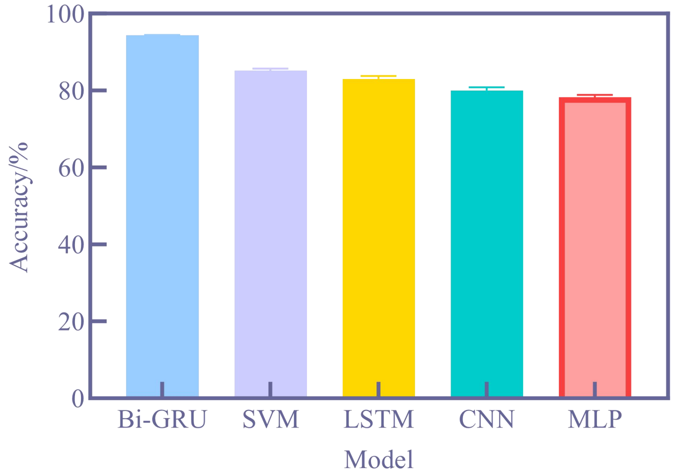


Figure 6. Accuracy of AI composing music melody and rhythm processing.

Seventy-six valid feature information were selected to compare the known music rules, and the AI composition music theory rule feature comparison is shown in **Table 1**. In the metric of notes out of tune, the score of this paper's method is 10.1%, which is the lowest among all methods, indicating that this paper's method is effective in maintaining the tonal accuracy of notes. In terms of average autocorrelation, the score of this paper's method is 0.14, which is slightly higher than other methods. Autocorrelation is a measure of similarity between elements in a sequence, generating a smoother or more predictable sequence of notes. In music analysis, higher autocorrelation may imply better coherence and harmony of the note sequences. For the indicator of notes out of tune, this paper's method significantly outperforms other methods with a 10.1% miss rate. This indicates that this paper's method has a significant advantage in maintaining the accuracy of notes in tune and is able to create more harmonized musical compositions Bidirectional LSTM networks, Convolutional Neural Networks and Multilayer Perceptual Machines scored 18.0%, 20.0%, and 17.0%, respectively, which are relatively high, implying that there is room for improvement in these methods in terms of maintaining the accuracy of notes in tune. In the assessment of interval difference less than an octave, the method of this paper achieved 70.0%, which is the best ranking among all the methods. Interval difference of less than an octave is an important measure of musical fluency and harmonization, and the high score of this paper's method indicates that it is able to maintain better musicality when dealing with interval relationships. On the metrics of having a unique maximum high note and having a unique minimum note, this paper's method achieved scores of 85.0% and 90.0%, respectively, and these two high scores may indicate that this paper's method is better able to maintain the uniqueness and richness of the notes when dealing with the range of the musical piece. The excellent performance demonstrated by this paper's method on several music evaluation metrics proves that the method has significant advantages and potentials in music processing and composition.

Table 1. Comparison of AI compositional music theory rule features.

Assessment indicators	Bi-GRU	SVM	LSTM	CNN	MLP
Excessive repetition of notes	63.30%	57.20%	58.30%	56.90%	48.30%
Average autocorrelation	0.14	0.12	0.1	0.08	0.11
Notes out of tune	10.10%	15.00%	18.00%	20.00%	17.00%
Intervals less than an octave apart	70.00%	65.00%	60.00%	55.00%	62.00%
Unique maximum high note	85.00%	80.00%	75.00%	70.00%	78.00%

Unique minimum note	90.00%	87.00%	83.00%	80.00%	85.00%

Table 1. *(Continued).*

4.3. Human aesthetic bias interactions

Figure 7 shows the scores of the music assessment indicators. The scoring criteria for various types of music are based on eight dimensions: melodic beauty, rhythmic coordination, tonal stability, innovative musical form, smooth rhythm, overall completeness, and listenability. Each dimension is quantitatively scored according to certain standards or expert evaluations, with a maximum score of 100 points for each dimension. Classical style music received high scores in a number of indicators, especially in tonality 95 and musical form 89, which reflected the deep heritage of classical music in terms of harmony, structural integrity and artistry. Meanwhile, its rhythm score of 85 and fluency score of 88 were also relatively high, showing the advantages of classical music in rhythm and melodic fluency. It shows that AI can well grasp the complexity and versatility of this music style. Pop music scored more balanced in all indicators, with a relatively high score of 84 for melody and 85 for completeness, indicating that pop music focuses on harmonious melody and clear structure to attract a wide range of listeners. Electronic Dance Music scored 86 in Music Form, indicating that AI had a good performance in dealing with the structure of this unique style of music. However, it scored low in Tonality 80 and may need to improve on the harmony of notes. Rock music performed better in terms of completeness 88 points and audibility 87 points, showing the advantages of rock music in terms of structural integrity and artistic aesthetics. However, its melody score of 86 and fluency score of 81 are relatively low, and it may need to pay more attention to melodic harmony and fluency. Jazz achieved high scores on most indicators, showing that AI-generated composed music in the jazz style has a high level in several aspects. The development of AI composition technology has had a significant impact on human music aesthetics. Firstly, the high level of AI compositions promotes the diversification and innovation of music styles, bringing a richer artistic experience to the human listener. The precision of AI compositions has also improved people's appreciation of musical details, and AI is able to accurately simulate the subtle characteristics of various musical styles, enabling listeners to be exposed to a more diverse range of musical works.

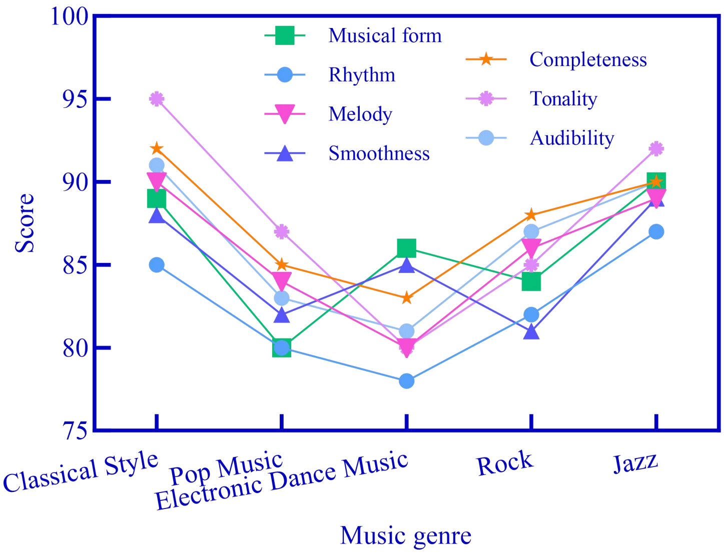


Figure 7. Music assessment indicator scores.

5. Conclusion

This study delves into AI composition technology and its impact on human music aesthetics. By using Bi GRU and self attention models, it is found that our music generation model performs well in simulating real music. The pitch of the generated notes is mainly concentrated between 55, 80, the duration of the notes is mainly concentrated below two beats, and the proportion of rest time is close to 78%. This is consistent with the situation in real music, indicating that the model can simulate the pitch distribution, duration, and body stop time of notes in real music very well. The music generation model achieved an accuracy rate of 94.28% in melody and rhythm data processing. The artificial intelligence composition model demonstrated the potential to imitate and generate different music styles, while addressing key aspects of human music aesthetics, particularly in jazz and classical styles, with scores of 92 and 90 respectively.

Although this study has achieved significant results, there are still some limitations, such as the need to improve the understanding and generation ability of the model enhancement model for complex harmonies, varied rhythms, and deep emotional expressions. Its applicability in real-world creative scenarios still needs further verification. In addition, the rapid development of artificial intelligence composition technology has had a profound impact on the music industry, musicians, and the creative process. Attention should be paid to the potential impact of this technology on musicians' professional identity, music copyright, and music aesthetic trends, ensuring that the healthy development of technology is balanced with humanistic concerns. Overall, this article not only deepens our understanding of the role of AI in music creation, but also addresses key issues in human music aesthetics, providing new perspectives and ideas for the deep integration of music and artificial intelligence in the future. Future research can further explore model optimization, interdisciplinary integration, and creative applications of AI composition, bringing more possibilities to the field of music creation.

Conflict of interest

The authors declare no conflict of interest.

References

1. Shi, N., & Wang, Y. (2020). Symmetry in computer-aided music composition system with social network analysis and artificial neural network methods. Journal of Ambient Intelligence and Humanized Computing, 1-16.
2. Zhu, H., Liu, Q., Yuan, N. J., Zhang, K., Zhou, G., & Chen, E. (2020). Pop music generation: From melody to multi-style arrangement. ACM Transactions on Knowledge Discovery from Data (TKDD), 14(5), 1-31.
3. Jin, H., & Yang, J. (2021). Using computer-aided design software in teaching environmental art design. Computer-Aided Design and Applications, 19(S1), 173-183.
4. Li, H. (2020). Piano automatic computer composition by deep learning and blockchain technology. IEEE Access, 8, 188951-188958.
5. Hong, J. W., Fischer, K., Ha, Y., & Zeng, Y. (2022). Human, I wrote a song for you: An experiment testing the influence of machines' attributes on the AI-composed music evaluation. Computers in Human Behavior, 131, 107239.
6. Pavlenko, O., Shcherbak, I., Viktoriia, H. U. R. A., Lihus, V., Maidaniuk, I., & Skoryk, T. (2022). Development of music education in virtual and extended reality. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 13(3), 308-319.
7. Gioti, A. M. (2020). From artificial to extended intelligence in music composition. Organised Sound, 25(1), 25-32.
8. Morreale, F. (2021). Where does the buck stop? Ethical and political issues with AI in music creation.
9. Pachet, F., Roy, P., & Carré, B. (2021). Assisted music creation with flow machines: towards new categories of new. Handbook of artificial intelligence for music: Foundations, advanced approaches, and developments for

creativity, 485-520.

10. Dai, D. D. (2021). Artificial intelligence technology assisted music teaching design. Scientific programming, 2021(1), 9141339.
11. Deruty, E., Grachten, M., Lattner, S., Nistal, J., & Aouameur, C. (2022). On the development and practice of ai technology for contemporary popular music production. Transactions of the International Society for Music Information Retrieval, 5(1), 35-50.
12. Briot, J. P. (2021). From artificial neural networks to deep learning for music generation: history, concepts and trends. Neural Computing and Applications, 33(1), 39-65.
13. Huertas-Abril, C. A. , & Palacios-Hidalgo, F. J. . (2023). Lgbtiq+ education for making teaching inclusive? voices of teachers from all around the world. Environment and Social Psychology.
14. Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Research, 6(4), 379-391.
15. Weng, S. S., & Chen, H. C. (2020). Exploring the role of deep learning technology in the sustainable development of the music production industry. Sustainability, 12(2), 625.
16. Shukla, S. (2023). Creative Computing and Harnessing the Power of Generative Artificial Intelligence. Journal Environmental Sciences And Technology, 2(1), 556-579.
17. Dash, A., & Agres, K. (2024). AI-Based Affective Music Generation Systems: A Review of Methods and Challenges. ACM Computing Surveys, 56(11), 1-34.
18. McCormack, J., Hutchings, P., Gifford, T., Yee-King, M., Llano, M. T., & D'inverno, M. (2020). Design considerations for real-time collaboration with creative artificial intelligence. Organised Sound, 25(1), 41-52.
19. Hernandez-Olivan, C., & Beltran, J. R. (2022). Music composition with deep learning: A review. Advances in speech and music technology: computational aspects and applications, 25-50.
20. Loughran, R., & O'Neill, M. (2020). Evolutionary music: applying evolutionary computation to the art of creating music. Genetic Programming and Evolvable Machines, 21, 55-85.
21. Ijiga, O. M., Idoko, I. P., Enyejo, L. A., Akoh, O., Ugbane, S. I., & Ibokette, A. I. (2024). Harmonizing the voices of AI: Exploring generative music models, voice cloning, and voice transfer for creative expression. World Journal of Advanced Engineering Technology and Sciences, 11(1), 372-394.