International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-16

https://doi.org/10.70917/ijcisim-2025-0268
Article

An Innovative Study on the Interaction Mode of Artificial Intelligence-Enabled Music Classroom

Wei Yao *

Music School, Taizhou University, Taizhou, Jiangsu, 225300, China; yaowei_sima@163.com

Abstract: AI music creation software can inspire students to create music, so that they can not only learn music theory and practical skills in the process of creating music, but also form a harmonious classroom interaction mode of teacher-student interaction, student-student interaction, and learning individual and teaching intermediary. In this regard, the study designed a music generation model based on Transformer-XL to generate higher quality music. The model consists of a fragment-level recursive mechanism and a new relative position encoding scheme, which is more conducive to improving the harmony of the generated samples. Comparative evaluation of this paper's model with some more advanced mainstream music generation models, such as MiDiNet and MusicVAE, in terms of objective metrics and subjective listening tests shows that the Transformer-XL model proposed in this paper outperforms other generative models at both the subjective and objective levels, and can generate higher quality music. The interactive effect of classroom based on AI music creation is evaluated based on the Flanders Interaction Analysis System, and the evaluation results reflect that AI music creation increases the interest and participation in the classroom, allows teachers and students to have a more diversified and interactive way of learning, and injects new vitality into the traditional music teaching, which makes the music classroom become more vivid and efficient.

Keywords: Transformer-XL; music generation; Flemish interaction; music classroom

1. Introduction

With the arrival of the intelligent era, artificial intelligence technology is changing our life and work style at an unprecedented speed [1]. The field of education, especially the field of music education, has also ushered in a profound change in this technological revolution [2]. Artificial intelligence technology not only enriches the teaching means, but also improves the teaching efficiency and injects new vitality into music education. Through intelligent teaching aids, personalized learning recommendations, intelligent music creation and other forms of application, AI empowers music education, innovates the teaching mode, improves the learning experience, and promotes the equalization of educational resources [3-4].

The innovation of music classroom interaction mode is driven by artificial intelligence, and the key lies in reconstructing the underlying logic of “teaching-learning” interaction [5]. Artificial intelligence, by leveraging natural language processing technology, analyzes students' voice and text feedback, captures real-time learning status data, and builds a dynamic student situation profile. This enables teachers to accurately grasp the cognitive bottlenecks of each student, achieving a transformation from "group indoctrination" to "individual response" [6-8]. Moreover, intelligent interactive tools, such as virtual teaching assistants and contextualized learning systems, can create immersive interactive scenarios, transform abstract music symbols into perceptible images and interactive tasks, and dissolve the problems of time and space constraints and expression thresholds in the traditional classroom [9-10]. The automated analysis function of artificial intelligence can quickly aggregate the generative resources in the interactive process to form a personalized learning path, which not only reduces the mechanical workload of teachers, but also improves the relevance and continuity of interaction by virtue of the closed loop of data feedback, so that classroom interactions move from “random triggering” to “precise design”

[11-13]. Artificial intelligence has brought new opportunities and challenges for music education, how to make full use of the advantages of artificial intelligence to empower music education has become an important topic of current research in the field of music education [14].

Intelligent technology has injected new interactive and interesting elements into the education of different disciplines, greatly stimulating students' interest and enthusiasm in learning, which has been confirmed on many studies [15-16]. Literature [17] used interactive digital comics teaching materials in nine elementary schools and proved the effectiveness of interactive digital comics teaching materials in improving the intensity of students' classroom participation and learning effectiveness through methods such as descriptive statistical analysis and one-way ANOVA. Literature [18] implemented an interactive teaching mode based on smart interactive tools in a neurobiology online course and compared the learning effects of this teaching mode with face-to-face teaching mode, and through descriptive statistical analysis, it was learned that the majority of the students believed that the smart interactive tools could help them to achieve their learning goals quickly. Literature [19] designed an interactive whiteboard and studied its application value in biology teaching. The flexibility and multifunctionality of the interactive whiteboard improved students' interaction and participation in the course, and had a positive effect on their motivation and knowledge understanding. Literature [20] systematically outlined the impact of digital storytelling in education on classroom interactivity, and the study concluded that the active learning environment built by the digital storytelling model for students promoted their online interactive learning and boosted their motivation and self-confidence. Literature [21] developed a teaching method called PollEverywhere Audience Response System, which is used in undergraduate dentistry courses to provide students with real-time teaching feedback, improve students' motivation and attention to learning, and at the same time, add a few points of fun and interactivity to the classroom. Literature [22] suggests that interactive learning technologies such as online discussions, virtual laboratories, and gamified activities can attract students' interest and participation in learning activities, and the combination of multimedia and interactive technologies can deepen students' understanding of complex concepts. Literature [23] evaluated the effectiveness of using photographic images as an online interactive teaching strategy, and during the study it was found that students are attracted to photographic images in the online classroom, which stimulates creative thinking, and the frequency of interactive behaviors among students in the online discussion community increases to provide students with an optimal learning experience.

In music education, some studies have found that interactive teaching of music based on artificial intelligence and computer technology can cultivate students' independent learning ability and enhance their motivation to learn music [24-25]. Literature [26] reveals the core role of interactive teaching strategies in music teaching, and carries out in-depth research on teacher-student interactive teaching strategies in the music classroom, and puts forward effective solutions to the problems in music teaching, providing a solid theoretical foundation for the innovation of music interactive teaching mode. Literature [27] proposes an interactive teaching music intelligence model based on the RBF algorithm, which consolidates the students' main position in the music classroom, improves the students' inquiry thinking, and provides a better interactive teaching strategy for music teaching. Literature [28] also used the interactive teaching music intelligence system based on RBF algorithm to obtain similar research results as above, in addition, it was found that the method improved the quality and efficiency of music education to a certain extent, and was able to assist students in learning music knowledge. Literature [29] used information and communication technology to design an interactive teaching model for high school music, in order to study the teaching effect of this model selected two groups of students for comparison, the results of the reality of interactive teaching model of all the students' music performance has been improved. Literature [30] explored the effectiveness of interactive teaching in music learning in terms of student engagement, creativity, cooperation, and critical thinking, and found that interactive teaching can promote active learning, accommodate diverse learning styles, and foster creativity. Literature [31] used virtual reality technology and artificial intelligence technology to create an interactive music course teaching mode, and the intelligent algorithm has a high recognition accuracy for low-frequency signals, medium-frequency signals and high-frequency signals, and its application to music teaching has a significant effect on students' learning ability. Literature [32] points out that high-quality interaction between teachers and students is necessary to achieve the goals of teacher supervision and student subjectivity, based on this, the use of computer-assisted systems proposed a teacher-student interaction music teaching strategy, and studied the effectiveness of this strategy. Literature [33] used pedagogical observations, content analysis of curriculum and educational platforms, and comparative analysis of digital tools' functions to confirm the effectiveness of digital interactive tools in music education, including the enhancement of musical skills such as aural and rhythmic skills.

However, the digital interactive education model does not completely replace the traditional teaching model in some subjects. Literature [34] designed an interactive teaching mode based on digital games to compare the teaching effectiveness with the traditional lecture mode in an elective course of radiology

medicine, and the experimental results negated the digital game teaching mode because the mode was not as effective as the traditional lecture mode in terms of test scores, learning interests, and so on. Moreover, there are some limitations and challenges in the application of digital intelligence technology in the field of education [35]. Literature [36] identified problems in the application of digital tools in educational work such as motivation to use digital tools, judging the validity of information sources, limitations due to national policies, plagiarism, filtering in search engines, and insensitivity to the online language, which can hinder the effective use of digital tools.

The study proposes a classroom interaction model based on AI music composition in order to enhance student classroom participation and improve the teacher-student interaction rate, and designs an automatic composition model for generating higher quality. The model is based on the Transformer network and proposes a stochastic mask model for composing, which solves the problem of contextual fragmentation by preprocessing the music data and designing a fragment-level recursive mechanism and relative position encoding to improve the harmony and richness of the generated music. The application of the model in classroom practice is analyzed based on the Flanders Interaction Analysis System with a view to improving the quality of AI-based music classroom teaching.

2. A study of classroom interaction based on AI music composition

The use of AI music creation software to assist teaching can increase students' attention to the classroom and interest in learning, so as to produce a positive response to the teacher's teaching, and students have the opportunity to go on stage to operate the APP and show their own music creation, which strengthens students' participation in the classroom, and enables more communication and interaction between teachers and students, and between students and students within the classroom. And students exposed to these software in the classroom can use the music APP to learn and entertain themselves after class, take the initiative to obtain music information, share music, discuss music, get experience, maintain the interest level in music, and also discuss music issues with teachers.AI music creation software into the music classroom, so that teachers and students have a more diversified and interactive way of learning, so that teachers students and other music learners, The transmitter creates synchronous or asynchronous communication and interaction, and the teacher no longer has absolute control over the knowledge, but stands on one side with the students and participates in the learning together, and makes a guide to the students' lifelong music learning.

In this chapter, the music generation model in AI music creation software will be investigated to design music models that can generate higher quality music models to be used to assist music teaching.

2.1. Classical Transformer Model

Early on, people used RNN [37] for compositing research, which always suffered from the problem of gradient vanishing when dealing with long sequences, and still struggled to establish long-distance dependencies. A research in machine translation made the Transformer model hot, and the Transformer network architecture is shown in Figure 1. Because of its breakthroughs in many domain tasks, its model architecture consists of an encoder and a decoder, where the encoder is stacked using a multi-head attention layer and a feed-forward layer, where both sub-layers use residual concatenation and normalization. The decoder module differs from the encoder in that it uses an attention mechanism with a mask that prevents the model from being trained is seeing future information [38].

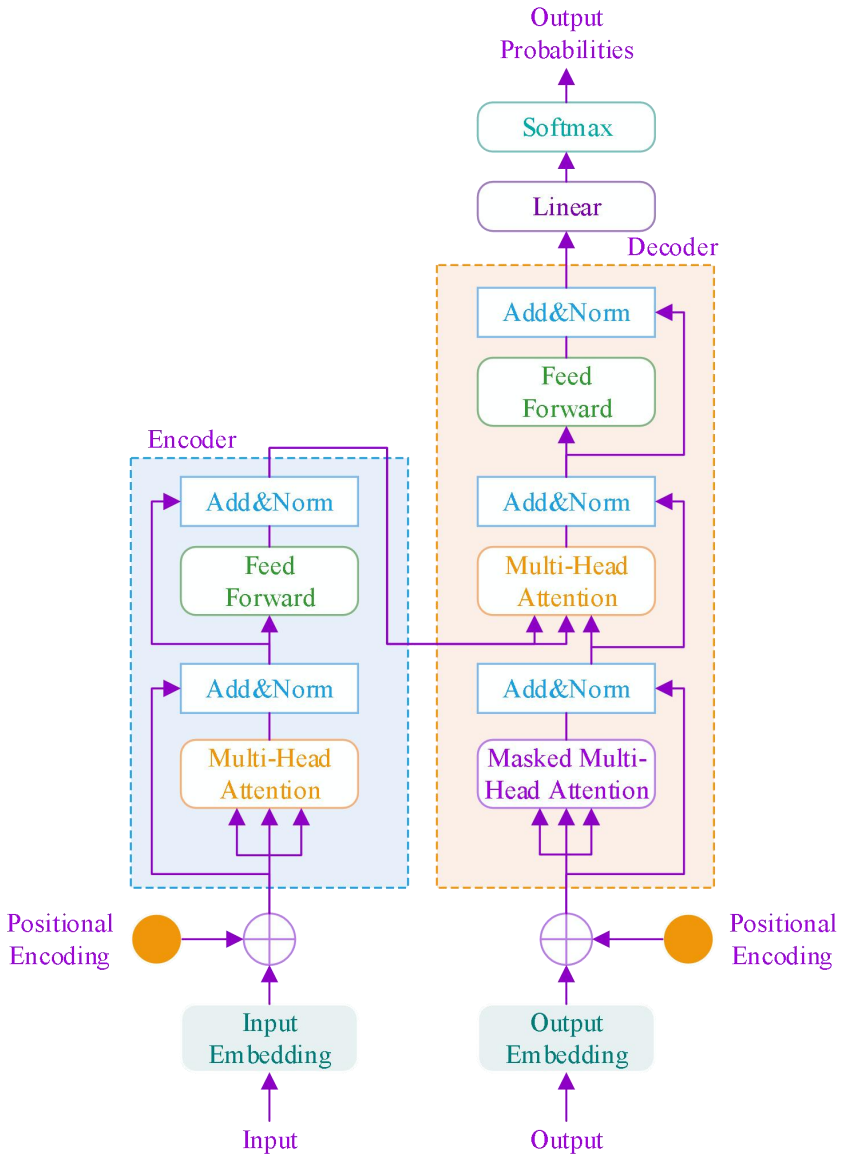


Figure 1. Transformer network architecture.

2.1.1. Multi-attention calculations

Attention calculation in Transformer uses the query-key-value approach. The computation process is shown in Figure 2. The input sequence $X=[x_1,\ldots,x_N]\in\mathbb{R}^{D_x\times N}, H=[h_1,\ldots,h_N]\in\mathbb{R}^{D_v\times N}$ are the sequence outputs, and the input sequences are each associated with the weight matrices $\mathbf{W_q}\in\mathbb{R}^{D_k\times D_x}, \mathbf{W_k}\in\mathbb{R}^{D_k\times D_x}, \mathbf{W_v}\in\mathbb{R}^{D_v\times D_x}$ performs dot product operation to obtain the query vector matrix $Q=[q_1,\ldots,q_N]$, key vector matrix $K=[k_1,\ldots,k_N]$, value vector matrix $\mathrm{V}=[v_1,\ldots,v_N]$, and mapping computation is Eq. (1), Eq. (2), Eq. (3), and each query vector can be Eq. (4) to get the attention of key pair values.

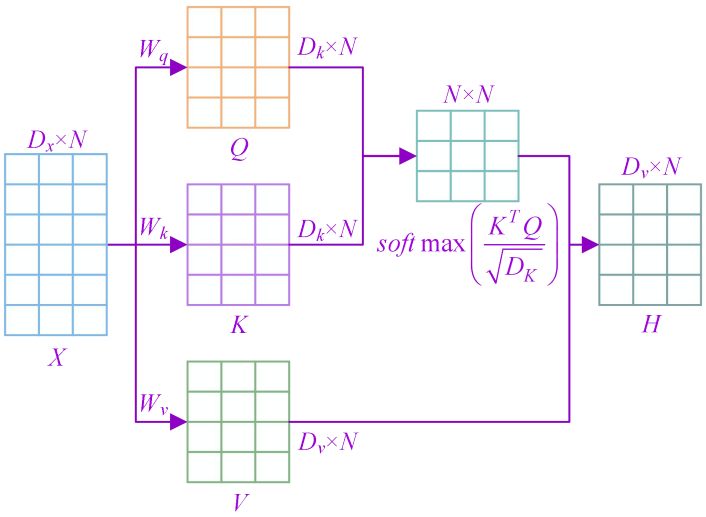


Figure 2. Self-attention computing process.

$$\mathbf{Q} = \mathbf{W_q X} \in \mathbb{R}^{D_k \times N} \tag{1}$$

$$\mathbf{K} = \mathbf{W_k X} \in \mathbb{R}^{D_k \times N} \tag{2}$$

$$\mathbf{V} = \mathbf{W_v X} \in \mathbb{R}^{D_v \times N} \tag{3}$$

$$Attention(Q,K,V) = soft\max\left(\frac{K^T Q}{\sqrt{D_k}}\right)V \tag{4}$$

Multi-head attention is to connect the output information of different heads as in Equation (5):

$$\begin{aligned} MultiHead(Q,K,V) &= Concat\left(head_1,\ldots,head_h\right)W^o \\ head_i &= Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \end{aligned} \tag{5}$$

2.1.2. Location coding

Because it is a purely attentional model, we cannot know the positional information of the sequence, so Transformer embeds positional coding in the bottom input of the encoder and decoder, there are many kinds of positional coding, here we introduce the absolute positional coding used in the original article of Transformer, which utilizes sine function and cosine function coding with different frequencies as in Eq. (6) and Eq. (7):

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_m}}\right) \tag{6}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_m}}\right) \tag{7}$$

pos is the position, $PE_{(pos,i)}$ is the value of that position of the position vector, and d_m is the dimension size of the mapping.

2.1.3. Feedforward Neural Networks

In the early days of neural network research, feed-forward neural networks were invented, which are relatively simple and straightforward network topologies. the position-based feed-forward neural network in Transformer is an all-connected feed-forward neural network same that will act on every position and finally activated using the ReLU function, see Equation (8):

$$FFN(x) = \max\left(0, xW_1 + b_1\right)W_2 + b_2 \tag{8}$$

where $W_1 \in \mathbb{R}^{d_m \times d_f}, W_2 \in \mathbb{R}^{d_f \times d_m}, b_1 \in \mathbb{R}^{d_f}, b_2 \in \mathbb{R}^{d_m}, d_f, d_m$ are the inner and input-output dimensions, respectively, and d_m is generally 502.

2.2. Music generation based on improved Transformer

2.2.1. Data pre-processing

Music preprocessing and network modeling are the two most important modules for music generation. In this section we will discuss note duration, pitch and chord content, we can process MIDI files to obtain this information and digitally represent it to obtain a multidimensional tensor, which after encoding and mapping can be directly fed into a neural network to learn. The main network architecture is based on Transformer's improved model, which utilizes a random masking mechanism.

Pitch is a concept closely related to frequency, as the frequency rises the pitch of the note rises. There are 88 keys on a traditional piano, including 52 white keys and 36 black keysThe piano range is from A0 to C8.In Western music there are twelve notes per octave: A, A#, B, C, C#, D, D#, E, F, F#, G, G#. The number following the name of the note indicates the octave in which the note is located, and all notes can be symbolically mapped. The number of the note we denote it as pitch, utilizing 88 integers to represent all piano key pitches.

The duration of a note is also an important feature of music. A beat is the unit of measure of the duration of a note; a whole note is 4 beats, and a 1/4 note is exactly one beat. A 1/4 note maps to the integer 1. In a MIDI-like model, the duration of a note must be inferred from the time interval between the opening of a note and the closing of the corresponding note; in this paper, we use the time interval directly as an element of the network input, which makes it easier to model musical melodies.

Chord Recognition: chord is another important component of music, so it is also used as an input event in this paper during preprocessing. A chord is a combination of notes that sound together in a particular order. A chord consists of a root note and several tones that are a few degrees above or below the root note. We use 12 chord root notes (A, A#, B, C, C#, D, D#, E, F, F#, G, G#), and in data preprocessing, we mainly carry out the recognition of major triads, diminished triads, augmented triads, and minor triads.

2.2.2. Modeling approach

(1) Transformer-XL.

Although the Transformer structure has the ability to learn long-term dependencies, its processing sequence length is fixed. If this length is exceeded, the sequence needs to be divided into multiple segments and trained independently. However, this processing prevents the model from learning any longer dependencies beyond the predefined length, and the fixed-length segmentation usually does not take semantic boundaries into account, which can lead to a model that lacks the necessary contextual information for better prediction.

To address this problem, this paper proposes the Transformer-XL model [39], which enables learning dependencies beyond the fixed-length limit without destroying temporal consistency. It consists of a fragment-level recursive mechanism and a new relative position encoding scheme. The study introduces the concept of recursion in deep self-attentive networks, and instead of computing the hidden states of each new segment from scratch, it reuses the hidden states obtained in previous segments and creates a cyclic connection between the segments by using the reused hidden states as memory for the current segment. The existence of this circular connection makes it possible to model long-term dependencies. Also, the problem of context fragmentation can be solved by passing the information from the previous segment. In addition, a relative positional encoding is proposed instead of absolute positional encoding in order to realize state reuse without causing temporal confusion.

(2) Fragment-level recursion mechanism

In order to solve the fixed-length context limitation, the authors introduce a recursion mechanism in the Transformer architecture. During training, the sequence of hidden states computed in the previous segment is cached and reused as an extended context when the model processes the next new segment. This additional input allows the network to utilize information from the history, thus enabling the model to capture long-term dependencies and avoid the context fragmentation problem.

Formally, two consecutive segments of length L are denoted as $S_\tau = \left[x_{\tau,1}, \cdots, x_{\tau,L}\right]$ and

$S_{\tau+1}=\left[x_{\tau+1,1},\cdots,x_{\tau+1,L}\right]$. Remembering that the hidden state generated by S_τ at the n th level is h_τ^n, the formula for the hidden state generated by $S_{\tau+1}$ at the n th level is:

$$\begin{aligned}\tilde{h}_{\tau+1}^{n-1}&=\left[SG\left(h_\tau^{n-1}\right)\circ h_{\tau+1}^{n-1}\right]\\ q_{\tau+1}^n,k_{\tau+1}^n,v_{\tau+1}^n&=h_{\tau+1}^{n-1}W_q^T,\tilde{h}_{\tau+1}^{n-1}W_k^T,\tilde{h}_{\tau+1}^{n-1}W_v^T\\ h_{\tau+1}^n&=Transformer-Layer\left(q_{\tau+1}^n,k_{\tau+1}^n,v_{\tau+1}^n\right)\end{aligned}\tag{9}$$

Where SG denotes that no gradient descent is involved in the computation, $\left[h_u\circ h_v\right]$ denotes that the two hidden sequences are concatenated by length dimension, and W denotes the model parameters. The key difference compared to Transformer is that the key $k_{\tau+1}^n$ and the value $v_{\tau+1}^n$ depend on the extension context $\tilde{h}_{\tau+1}^{n-1}$, and thus h_τ^{n-1} is cached from the previous fragment,. By applying this recursive mechanism between every two consecutive segments, it allows the model to essentially implement segment-level recursion in the hidden state. As a result, the effective context utilized by the model extends well beyond two segments.

(3) Relative position encoding

After the introduction of the recursive mechanism, the continuation of the original Transformer's position encoding approach would result in the same absolute position encoding for each fragment, making the network unable to correctly capture the position information of the sequence. Therefore, the authors propose a new relative position encoding method, which no longer considers the absolute position of a word in a fragment, but encodes the relative position between two words, and adds the relative position encoding to the internal of self-attention calculation, so that the model processes the position information from the network structure.

Specifically, the original absolute position coding and relative position coding formulas are shown in Eq. (10) and Eq. (11), respectively:

$$A_{i,j}^{abs}=E_{x_i}^TW_q^TW_kE_{x_j}+E_{x_i}^TW_q^TW_kU_j+U_i^TW_q^TW_kE_{x_j}+U_i^TW_q^TW_kU_j\tag{10}$$

$$A_{i,j}^{rel}=E_{x_i}^TW_q^TW_{k,E}E_{x_j}+E_{x_i}^TW_q^TW_{k,R}R_{i-j}+u^TW_{k,E}E_{x_j}+v^TW_{k,R}R_{i-j}\tag{11}$$

Compared to Transformer's absolute position encoding, Transformer-XL makes several main changes. First, the absolute position encoding U is replaced by the relative position encoding R_{i-j}, which is the same as that of Transformer, R is a sinusoidal encoding matrix with no learnable parameters. Second, two learnable parameters $u\in R^d$ and $v\in R^d$ are introduced to replace the query vectors $U_i^TW_q^T$ in the Transformer, showing that the query vectors corresponding to all query positions are the same, i.e., regardless of the query position, the attentional bias remains consistent across words. Finally, W_k is split into $W_{k,E}$ and $W_{k,R}$ to generate content-based key vectors and position-based key vectors, respectively. The relative position coding is then embedded into the Transformer-XL complete architecture after the recursive mechanism is computed as described below, where $n=1,\ldots,N,N$ is the number of layers of the network:

$$\begin{aligned}\tilde{h}_\tau^{n-1}&=\left[SG\left(m_\tau^{n-1}\right)\circ h_\tau^{n-1}\right]\\ q_\tau^n,k_\tau^n,v_\tau^n&=h_\tau^{n-1}W_q^{nT},\tilde{h}_\tau^{n-1}W_{k,E}^{n\ T},\tilde{h}_\tau^{n-1}W_v^{nT}\\ A_{\tau,i,j}^n&=q_{\tau,i}^{n\ T}k_{\tau,j}^n+q_{\tau,i}^{n\ T}W_{k,R}^nR_{i-j}+u^Tk_{\tau,j}+v^TW_{k,R}^nR_{i-j}\\ a_\tau^n&=Masked-\mathrm{Softmax}\left(A_\tau^n\right)v_\tau^n\\ o_\tau^n&=LayerNorm\left(Linear\left(a_\tau^n\right)+h_\tau^{n-1}\right)\\ h_\tau^n&=Positionwise-Feed-Forward\left(o_\tau^n\right)\end{aligned}\tag{12}$$

2.3. Design of Classroom Interaction Model Based on AI Music Composition

In the music classroom, allowing students to freely choose their favorite music styles and musical characteristics, and using the music generator based on the improvement of Transformer to randomly generate a complete musical work, including melody, harmony and lyrics, can bring a new interactive experience to the music classroom. This innovative teaching method can not only improve students' participation in the music classroom, but also enable them to intuitively feel the contrast between different music styles, which provides a precursor to the teaching of music style appreciation in the following. The easy-to-understand operation mode of the AI music creation model also opens up the idea of music creation for music learners who have no foundation in composition. By entering a simple command, such as "a melancholy blues", the system can automatically generate a piece of music that meets the characteristics of the command. This low-barrier way of music creation is especially suitable for stimulating students' desire to explore musical styles in music appreciation classes. In the music classroom, teachers can create a music theme and then divide students into several groups to practice AI music creation. Students can design their own music style and specific features through group collaboration, and finally share the results with other groups. This teaching mode of "creation-experience-interaction" can effectively change the limitations of one-way listening of students in traditional music appreciation class, which can not only enliven the classroom atmosphere, but also allow students to deepen their understanding and feelings of music in personal participation.

3. Generating music and assessing classroom interactivity

This chapter provides a subjective and objective assessment of the music generator constructed above, and explores the impact of the classroom interaction model of AI music composition on classroom interaction outcomes based on teaching practice.

3.1. Objective assessment of generated music

For the selection of objective assessment indicators, for pitch, this paper uses the number of notes (X1), the number of note classes (X2), pitch consistency (X3) and pitch class entropy (X4). For rhythm, this paper uses air beat rate (X5) and rhythmic interval similarity (X6).

Number of notes: the number of individual notes used in a song, with an upper limit of 132, by analyzing the number of notes a reasonable judgment can be made on the interval of pitch.

Number of note classes: the number of note classes used in a song, with an upper limit of 15. For different music, the use of different note classes can make the music richer and more complete.

Pitch Consistency: The maximum pitch rate on all major and minor scales in a song, which can be calculated using the fraction of pitches in the standard scale, higher pitch consistency indicates a higher pitch rate in a song.

Pitch class entropy: is defined as the entropy value of the normalized pitch class histogram. In information theory, entropy is a measure of the "uncertainty" of a probability distribution, and is used here as a metric to help assess the tonal quality of music, i.e.:

$$\text{pitch\_class\_entropy} = -\sum_{i=0}^{11} P(\text{pitch class} = i) \times \log_2 P(\text{pitch class} = i) \tag{13}$$

In Eq. (13), i denotes the 12 note categories, and when a song has a smaller entropy value and a denser use of scales, it indicates a more stable pitch.

Empty beat rate: the ratio of notes not played to the total number of beats, it can easily appear as a disturbance term in the processing of the generated dataset.

Rhythmic interval similarity: the average Hamming distance (Hamming distance: the number of different notes in corresponding positions in two measures) of neighboring measures, i.e:

$$\text{groove\_consistency} = 1 - \frac{1}{B-1} \sum_{i=1}^{B-1} d\left(G_i, G_{i+1}\right) \tag{14}$$

In the formula B is the number of bars in the song, which represents the binary start vector of the G_i th i bar and is the Hamming distance between G_i and G_{i+1}. It can help us to evaluate how good the rhythm of a song is; a better rhythm of a song means that it has a higher similarity of rhythmic intervals.

In this paper, three checkpoints with losses equal to 0.1, 0.4, and 0.7 are selected for objective metrics evaluation of the music generated by the Transformer model as well as Transformer-XL.

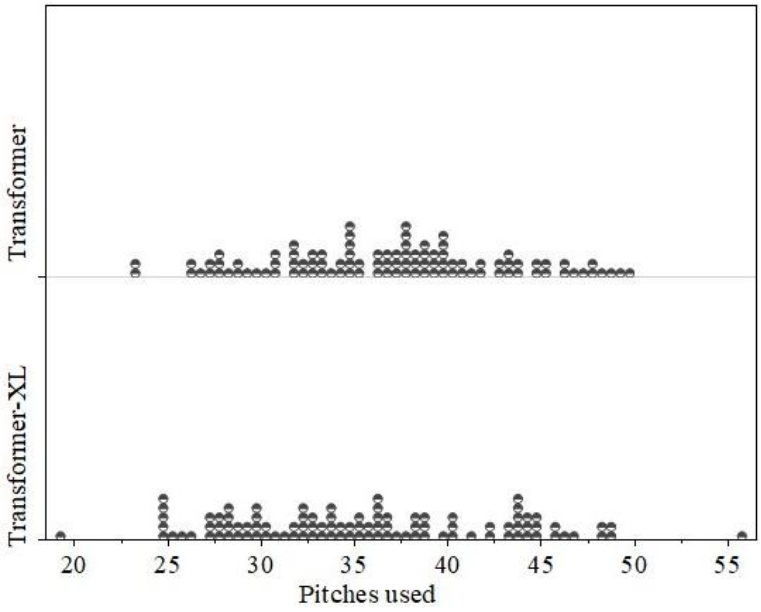
The experimental steps are: in this paper, 20 32-bar MIDIs generated by the two models are selected at each checkpoint, and in addition, 600 MIDIs are selected from the database, and the evaluation

experiments are conducted on these 600 MIDIs by utilizing the six evaluation indexes mentioned above. The comparison results of the evaluation metrics are shown in Table 1. The music generated using Transformer-XL has higher scale consistency (X3) and rhythmic interval similarity (X6) as well as lower pitch class entropy (X4) compared to the original Transformer, which suggests that the music generated using Transformer-XL has better rhythmic and melodic properties. When the loss is equal to 0.1, the scale consistency, rhythmic interval similarity and pitch class entropy of Transformer-XL are 0.964, 0.994, and 2.721, respectively.

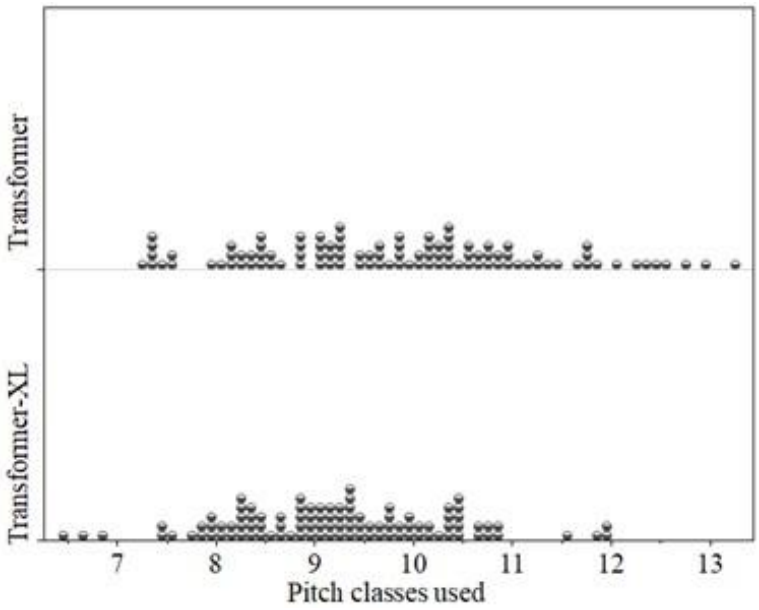
Table 1. The comparison results of the evaluation index.

Index	Model	Loss		
		0.7	0.4	0.1
X1	Transformer	35	34	35
	Transformer-XL	38	37	35
X2	Transformer	8	10	8
	Transformer-XL	8	10	8
X3	Transformer	0.922	0.956	0.943
	Transformer-XL	0.926	0.973	0.964
X4	Transformer	2.946	2.741	2.736
	Transformer-XL	2.933	2.723	2.721
X5	Transformer	0.36	0.36	0.14
	Transformer-XL	0.42	0.41	0.13
X6	Transformer	0.963	0.926	0.971
	Transformer-XL	0.967	0.934	0.994

The number of notes and the number of note classes of the 100 songs involved in the validation of each model are shown in Fig. 3, with (a) and (b) denoting the number of notes and the number of note classes, respectively. It can be found that the music generated using the improved Transformer is more stable in terms of the number of notes and the number of note classes compared to the original Transformer, and does not generate songs with large variations, which reflects the fact that Transformer-XL has greater stability and fewer error rates.



(a) Pitches used



(b) Pitch classes used

Figure 3. The number of notes and the number of notes.

In order to verify the quality and performance of the Transformer-XL model more comprehensively, this paper compares the MiDiNet model based on generative adversarial networks and the MusicVAE model based on a two-layer encoder using the six objective evaluation metrics mentioned above.The MiDiNet model and the MusicVAE model each generate 50 pieces of 32-bar music, which are later compared with the Transformer-XL model generated by the Transformer-XL model with a loss of 0.1 The results of the evaluation comparison are shown in Table 2. The comparison shows that using Transformer-XL is much better than MiDiNet and MusicVAE in the evaluation of the generated music, which verifies that using the Transformer-XL model has a certain degree of disparity in the task of music generation compared to the use of Generative Adversarial Networks and Bilayer Encoders, which are based on generative adversarial networks. Transformer-XL model for music generation has a clear advantage.

Table 2. Evaluate the results.

Index	Transformer-XL	MiDiNet	MusicVAE
X1	35	34	32
X2	8	7	6
X3	0.964	0.961	0.943
X4	2.721	3.641	3.216
X5	0.13	0.63	0.89
X6	0.994	0.923	0.906

3.2. Subjective assessment of generated music

The experiment invited 20 professional musicians to perform aural evaluations of the generated scores. The scores were derived from music from samples of tracks 1, 2, 3, and 4 generated by four automated composition models and from the training concentrator's work piece. A total of ten pieces of music were randomly selected from each of them and given to each listener to be scored after listening. In order to avoid aesthetic fatigue of the subjects, each piece of music was played with a fixed clip of 15 seconds.

The listening evaluation adopts the rating scoring method, which is covered in the current domestic and international sound or music standards, for example, there is the national standard “Methods and Technical Requirements for Subjective Evaluation of Sound in Radio Programs”. There are some differences in the specific settings of the methods in the above standards, but all of them can carry out

subjective grade scoring evaluation on sound quality, music listening sense and other aspects.

This section adopts the current national standard GB/T 16463-1996 rating scoring method, and its scoring level is shown in Table 3. "1~5" for the grade "bad, poor, medium, good, excellent".

Table 3. Music level rating table.

Score	Grade
5	Excellent
4	Good
3	Middle
2	Bad
1	Unwell

The scores were summarized based on the music of each compositional method, showing the average, highest, and lowest scores as shown in Figure 4. As far as the manual subjective evaluation is concerned, the results show that among the automatic composition models, Transformer-XL has the highest average score (4.06), and Transformer (3.63) is in the second place. On the other hand, neither of the two automatic composition models based on generative adversarial networks and bilayer encoders scored very high. Overall there is still a gap between the automatic composition models and the human compositions, but as far as the scores (highest scores) of the individual compositions are concerned, the automatic composition models are still able to achieve good results on the human subjective evaluation.

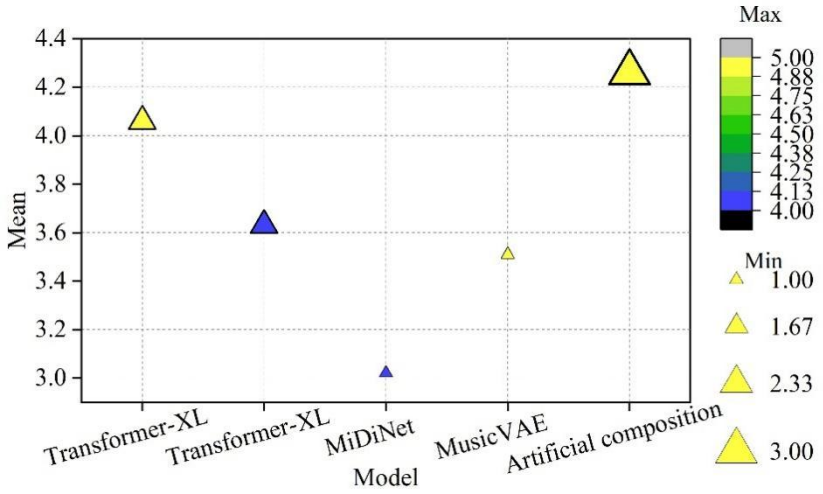


Figure 4. Artificial evaluation scores of different composing methods.

3.3. Assessment of classroom interactivity

The study selected the Transformer-XL model designed in this paper for music composition, and selected 60 students from the sophomore (1) class of music majors in School A as the research subjects to empirically analyze the classroom interaction mode based on AI music composition based on the Flanders interaction analysis method.

The Flanders interaction analysis system is roughly composed of three parts: (1) Codes describing classroom behaviors. (2) Observation and recording of codes. (3) Counting the numbers of behaviors. The basic method of the Flanders Interaction Analysis System is to classify classroom teacher-student interaction behaviors into three main categories of 10 situations. Among them, teachers' verbal behavior accounts for 7 situations, students' verbal behavior is divided into 2 situations, and another invalid verbal behavior is set. The specific content is shown in Table 4.

Table 4. Flanders interactive analysis system classification.

Classification		Code	Content
Teacher language (D1)	Indirect influence	1	Express emotion
		2	Question
		3	Praise
		4	Adopt opinions
	Direct influence	5	Teach
		6	Instruction
		7	Criticize
Student language (D2)	Passive language	8	Answer questions
	Active language	9	Active speech
Silencing (D3)		10	Speechless

How to reflect the teacher's leading role in teaching and effectively implement the teaching and learning process. How teachers accomplish their teaching tasks relies on having good teaching literacy and pedagogy, being able to effectively guide students through various teaching activities, being able to pay attention to different individuals so as to form an effective teaching guidance, and being able to make meaningful assessment of students' learning so as to improve the effectiveness of teaching. In response to these requirements, the content of the record of teachers' teaching situation produced is shown in Table 5. Four evaluation dimensions of presentation, instruction, evaluation and strain were included.

Table 5. Teacher teaching records content.

Classification	Code	Content
Rendering (D4)	1	Plate
	2	Media presentation
Guide (D5)	3	Guide student learning
	4	Students take the stage
	5	Difficult students
	6	Cooperative learning
Evaluation (D6)	7	Process evaluation
	8	Knowledge (Works) evaluation
Strain (D7)	9	Adjust content
	10	Adjust the teaching process

The statistics of teachers' and students' behaviors in the classroom of music composition based on the Transformer-XL model are shown in Figure 5. It can be seen that:

(1) The teacher's language ratio, student's language ratio, and teacher's guidance ratio are 16.3%, 25.9%, and 19.2% respectively. It shows that this class implements the teaching concept of student-oriented and teacher-led, and the students' language is mainly reflected in group cooperative learning and answering questions. The teacher's questions are mainly closed-ended and a small part is open-ended, which develops students' language expression ability. In group cooperation, students can actively express their creative opinions and communicate with others, reflecting the teaching methods of inquiry learning and student independent learning.

(2) The ratio of silence is 28.7%, which is mainly based on students' independent operation and practice, watching their classmates' operation, etc. There is no ineffective language, and the classroom efficiency is high, which is in line with the observation.

(3) The content of teaching evaluation is 4.8%. According to the teaching observation, the evaluation

methods used by the teachers are process evaluation and evaluation of music creation works, and the evaluation methods are diversified through students' operation on the stage and group evaluation.

(4) The strain ratio point is 0.8%, which indicates that there is sufficient preparation before the lesson and most of the students are able to complete the learning tasks. In teaching, the teacher is able to improve students' learning efficiency by using individual counseling and individual tasks for individual students with learning difficulties.

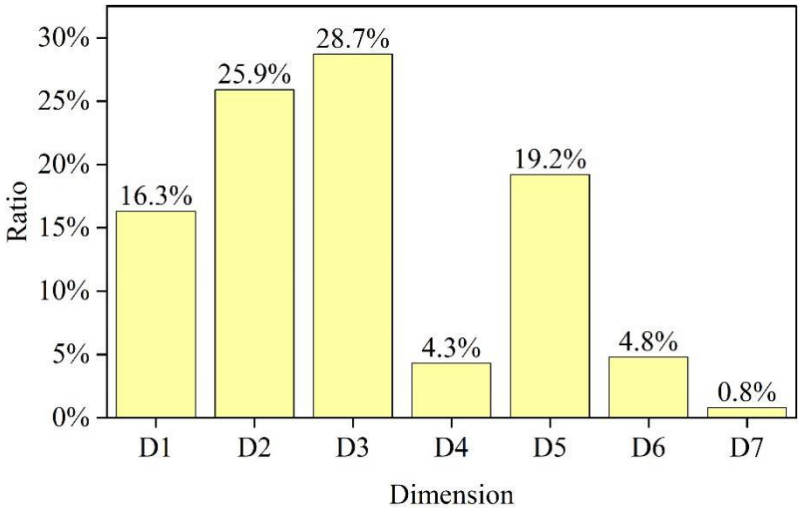


Figure 5. The rate of classroom teaching behavior.

In order to analyze the interactivity of the teacher-student classroom in a more systematic and detailed way, the dynamic features drawn according to the teacher-student classroom language ratio are shown in Figure 6. The horizontal coordinate of the dynamic characteristic curve indicates time in minutes, in the example classroom is 40 minutes, so the range of the horizontal coordinate is from 0 to 40 minutes. The vertical coordinate indicates the teacher/student language percentage. The solid line in the table indicates the teacher language dynamic ratio and the underlined line indicates the student language ratio. Observing Figure 6, both curves have relatively large amplitude of vibration. The teacher language dynamic ratio has little amplitude in the first half of the lesson (before the 22nd minute), which is basically close to or more than 90%, but it changes relatively dramatically in the second half of the lesson (after the 22nd minute), and the curve basically stays below 50% after the 22nd minute and before the 36th minute, and then rises rapidly after the 36th minute. Compared to the teacher language ratio curve, the student language ratio curve was essentially the opposite, with the student language ratio remaining at a relatively low value during the first half of the lesson, and then increasing significantly after the 22nd minute. This is mainly due to the music classroom model based on AI music composition, the first half of the lesson is the teacher's explanation of the requirements of the composition, mainly in the form of teacher's lecture and Q&A interaction between the teacher and students, and then enter the stage of student communication and interaction tasks, the students' language ratio increased rapidly, and the teacher's feedback on the results of the music composition after the end of the interactive task, which is a more frequent interaction in this stage, but the overall teacher evaluation is still dominated by the teacher. In this stage, the interaction is more frequent, but the teacher's evaluation is still the main focus. In the second half of the classroom, it can be seen that the two curves of the teacher and the students intersect with each other. It can be seen that the teacher and students interacted with each other frequently and well in the classroom. Every time the teacher raised a question, the students answered it, and the students also gave feedback to the teacher, so there was a back and forth between the teacher and students, and the classroom always maintained a warm atmosphere.

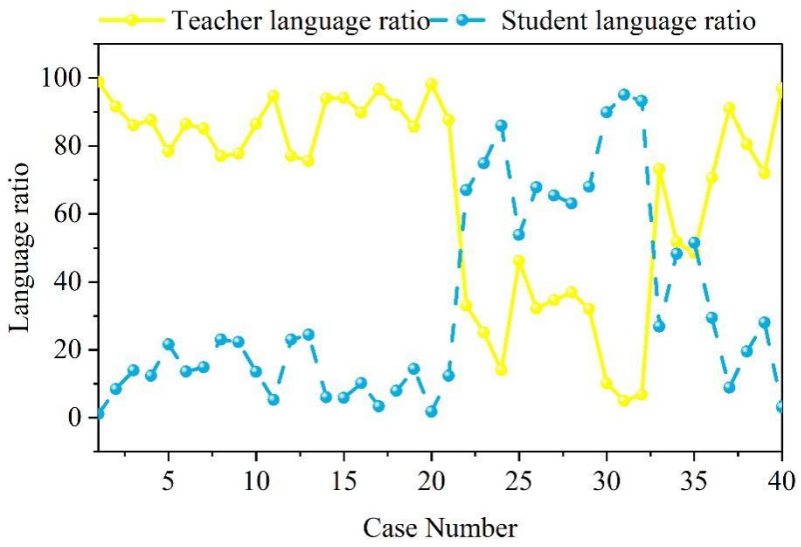


Figure 6. The dynamic characteristics of the language ratio of teachers and students.

4. Conclusion

The study proposes a classroom interaction model based on AI music composition and designs a music generation model based on an improved Transformer. The quality of the generated music is comprehensively evaluated using a combination of objective and subjective evaluation. Experimental results show that the improved model generates music with better rhythmicity compared to the original model as well as mainstream music generation models. The model was evaluated for its effectiveness in teaching practice using the Flanders Interactive Analysis System. The teacher language ratio and student language ratio were 16.3% and 25.9%, respectively, and the teacher and students interacted frequently in the classroom, always maintaining an enthusiastic classroom atmosphere. It shows that the classroom interaction model based on AI music composition can realize teacher-led and student-led in order to make the optimal integration of teaching and learning.

References

1. Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. Science, 380(6650), 1108-1109.
2. Jia, F., Sun, D., & Looi, C. K. (2024). Artificial intelligence in science education (2013–2023): Research trends in ten years. Journal of Science Education and Technology, 33(1), 94-117.
3. Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. British Journal of Educational Technology, 50(6), 2824-2838.
4. Relmasira, S. C., Lai, Y. C., & Donaldson, J. P. (2023). Fostering AI literacy in elementary science, technology, engineering, art, and mathematics (STEAM) education in the age of generative AI. Sustainability, 15(18), 13595.
5. Fang, H., Shu, L., Wang, X., & Hong, X. (2024, September). Research on Intelligent Classroom Interaction Analysis and Its Data Fusion Model. In 2024 4th International Conference on Educational Technology (ICET) (pp. 568-572). IEEE.
6. Kehl, K. L., Elmarakeby, H., Nishino, M., Van Allen, E. M., Lepisto, E. M., Hassett, M. J., ... & Schrag, D. (2019). Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. JAMA oncology, 5(10), 1421-1429.
7. Alhawiti, K. M. (2014). Natural language processing and its use in education. International Journal of Advanced Computer Science and Applications, 5(12), 72-76.
8. Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023). A systematic literature review on the applications of robots and natural language processing in education. Electronics, 12(13), 2864.
9. Bokhari, M. U., Ahmad, S., Alam, S., & Masoodi, F. (2011). Modern tools and technologies for interactive learning. environment, 13(15), 17-18.
10. Hina, S., Dominic, P. D. D., & Zaidi, K. S. (2020). Use of interactive tools for teaching and learning practices in higher education institutions. International Journal of Business Innovation and Research, 22(4), 469-487.
11. Kanchon, M. K. H., Sadman, M., Nabila, K. F., Tarannum, R., & Khan, R. (2024). Enhancing personalized learning: AI-driven identification of learning styles and content modification strategies. International Journal of Cognitive Computing in Engineering, 5, 269-278.

12. Van den Hurk, H. T. G., Houtveen, A. A. M., Van de Grift, W. J. C. M., & Cras, D. W. P. (2014). Data-feedback in teacher training. Using observational data to improve student teachers' reading instruction. Studies in Educational Evaluation, 42, 71-78.
13. Di Paola, F., Pedone, P., & Pizzurro, M. R. (2013). Digital and interactive Learning and Teaching methods in descriptive Geometry. Procedia-Social and Behavioral Sciences, 106, 873-885.
14. Cui, X., & Chen, M. (2024). A novel learning framework for vocal music education: An exploration of convolutional neural networks and pluralistic learning approaches. Soft Computing-A Fusion of Foundations, Methodologies & Applications, 28(4).
15. Papadimitriou, S., & Virvou, M. (2025). Computer Games for Entertainment and Education: A Literature Review and Exploration on Artificial Intelligence Integration. Artificial Intelligence—Based Games as Novel Holistic Educational Environments to Teach 21st Century Skills, 25-62.
16. Prananta, A. W., Rohman, A., Agustin, R., & Pranoto, N. W. (2024). Augmented reality for interactive, innovative and fun science learning: Systematic literature review. Jurnal Penelitian Pendidikan IPA, 10(SpecialIssue), 45-51.
17. Khotimah, H., & Hidayat, N. (2022). Interactive digital comic teaching materials to increase student engagement and learning outcomes. International Journal of Elementary Education, 6(2), 245-258.
18. Wang, P., Ma, T., Liu, L. B., Shang, C., An, P., & Xue, Y. X. (2021). A comparison of the effectiveness of online instructional strategies optimized with smart interactive tools versus traditional teaching for postgraduate students. Frontiers in psychology, 12, 747719.
19. Yang, K. T., & Wang, T. H. (2012). Interactive white board: Effective interactive teaching strategy designs for biology teaching. Tech, E-learning-engineering, on-job training and interactive teaching, 139-154.
20. Rajendran, V., & Yunus, M. M. (2021). Interactive learning via digital storytelling in teaching and learning. International Journal of Education and Literacy Studies, 9(3), 78-84.
21. Abdel Meguid, E., & Collins, M. (2017). Students' perceptions of lecturing approaches: traditional versus interactive teaching. Advances in medical education and practice, 229-241.
22. Adeoye, M. A., & Akinnubi, O. P. (2023). Integrating interactive learning technologies into traditional teaching methods for private higher education institutions. Formosa Journal of Computer and Information Science, 2(2), 223-234.
23. Perry, B. (2006). Using photographic images as an interactive online teaching strategy. The Internet and Higher Education, 9(3), 229-240.
24. Webster, P. R. (2017). Computer-based technology and music teaching and learning. In Critical essays in music education (pp. 321-344). Routledge.
25. Merchán Sánchez-Jara, J. F., González Gutiérrez, S., Cruz Rodríguez, J., & Syroyid Syroyid, B. (2024). Artificial intelligence-assisted music education: A critical synthesis of challenges and opportunities. Education Sciences, 14(11), 1171.
26. Chen, Y. (2024). Review of Interactive Teaching Methods of Music in Junior Middle Schools. Asian Journal of Education and Social Studies, 50(2), 49-59.
27. Chen, X. (2021, December). Research and application of interactive teaching music intelligent system based on artificial intelligence. In International Conference on Artificial Intelligence, Virtual Reality, and Visualization (AIVRV 2021) (Vol. 12153, p. 1215302). SPIE.
28. Liu, C., & Li, S. (2023, June). Design of Interactive Teaching Music Intelligent System Based on AI and Big Data Analysis. In International Conference on Computational Finance and Business Analytics (pp. 333-341). Cham: Springer Nature Switzerland.
29. Chao-Fernandez, R., Román-García, S., & Chao-Fernandez, A. (2017). Analysis of the use of ICT through music interactive games as educational strategy. Procedia-Social and Behavioral Sciences, 237, 576-580.
30. Sultanov, A. (2025). Educational Effectiveness of Using Interactive Methods in Music Pedagogy. Academic Journal of Science, Technology and Education, 1(1), 9-12.
31. Yan, J., & Xia, X. (2024). Interactive audio-visual course teaching of music education based on VR and AI support. International Journal of Human–Computer Interaction, 40(13), 3552-3559.
32. Huang, Y. (2022). Teacher-Student Interactive Creation Strategies in Music Teaching Assisted by Computer Information Technology. Mathematical Problems in Engineering, 2022(1), 5443729.
33. Aliksiichuk, O., Borysova, T., Kartashova, Z., Priadko, O., Kuziv, M., & Chaban-Chaika, S. (2025). Modern Digital Approaches to Training Music Teachers: Evolution from Classical to Interactive. International Journal on Culture, History, and Religion, 7(SI1), 273-296.
34. Courtier, J., Webb, E. M., Phelps, A. S., & Naeger, D. M. (2016). Assessing the learning potential of an interactive digital game versus an interactive-style didactic lecture: the continued importance of didactic teaching in medical student education. Pediatric Radiology, 46(13), 1787-1796.
35. Qiao, W., & Fu, J. (2023). Challenges of engineering education in digital intelligence era. Journal of Educational Technology Development and Exchange (JETDE), 16(2), 145-159.
36. Mısır, H. (2018). DIGITAL LITERACIES AND INTERACTIVE MULTIMEDIAENHANCED TOOLS FOR LANGUAGE TEACHING AND LEARNING. International Online Journal of Education & Teaching, 5(3).
37. Mohit Dua,Rohit Yadav, Divya Mamgai & Sonali Brodiya. (2020). An Improved RNN-LSTM based Novel Approach for Sheet Music Generation. Procedia Computer Science,171(C),465-474.https://doi.org/10.1016/j.procs.2020.04.049.
38. Yifei Zhang. (2025). An IoT-enhanced automatic music composition system integrating audio-visual learning with transformer and SketchVAE. Alexandria Engineering Journal,113,378-390.https://doi.org/10.1016/J.AEJ.2024.10.115.

39. Yunze Liang, Halidanmu Abudukelimu, Jishang Chen, Abudukelimu Abulizi & Wenqiang Guo. (2025). MAML-XL: a symbolic music generation method based on meta-learning and Transformer-XL. Multimedia Systems,31(3),206-206.https://doi.org/10.1007/S00530-025-01803-8.