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

Optimization of computer-assisted teaching strategies in music education: a case study of preschool music theory teaching

Peijun Xie[1,*]

[1] Ankang Vocational and Technical College, Ankang, Shaanxi, 725000, China

SUMMARY: *Music development cannot be detached from the development of computer technology, which will not only affect the music creation and performance process but will also inevitably have an impact on the development of means for music theory education. Based on the four focuses of music theory knowledge, the realization of computer music data expression through MIDI, the establishment of the music generation model by integrating RNN, GAN, and VAE, and saving the four-dimensional music information matrix in the music dataset, the MIDI file data feature extraction is conducted to get the monophonic and chord sequences of all musical instrument data sets, and the optimization of MIDI music generation models by WaveNet predictions. By analyzing the automatic composition results of music model RNN-VAE-GAN by using objective criteria, detecting the pitch variation of the main melody produced by RNN-VAE-GAN, and measuring the consistency of the note pitch variation of the MIDI music monophony composed by the model with the pitch variation path of the notes. According to the test results combining pre- and post-test and control experiments, comparing the effect of music learning between the experimental group and the control group, P-value of the core music literacy of the two groups was 0.005, and P-value of music learning interest of the two groups was 0.001, indicating a significant difference, which demonstrates a high level of progress in computer-assisted pre-school music theory teaching model application practice.*

KEYWORDS: *MIDI; WaveNet; RNN-VAE-GAN; music theoretical knowledge; preschool music theory teaching*

1 Introduction

Since pre-school education is the initial stage of individual “lifelong learning”, quality pre-school education will play a positive role in promoting the personality, physical and mental health, moral character, language expression, and “three views” of the early childhood group [1-4]. In this regard, it is important to further improve the quality of preschool education. In this regard, it is of great practical significance to further improve the overall quality of teaching in the field of preschool education. Therefore, the music curriculum in kindergarten can effectively stimulate the individual senses, brain and spiritual world of young children through music rhythm, melody, rhythm, lyrics and other detailed factors, and gradually cultivate their desire to know, explore and learn music, and then improve the comprehensive quality of young children [5-9]. Therefore, it is important to pay attention to preschool music education, and deeply and correctly recognize that preschool music education plays an important role in the healthy growth of young children. Among them,

*soulyoung@tzc.edu.cn
https://doi.org/10.65102/is2026494 Published: 30 April, 2026

music theory as an important part of music education, its teaching content is abstract, children are difficult to understand, resulting in low interest in learning, which in turn does not play a role in healthy growth [10, 11].

Computer technology plays an important role in the teaching process of music education, both in terms of the educational concept, educational content or even the curriculum [12]. As the emphasis on education increases, computer technology-aided teaching practices have now become the biggest research hotspot in the field of education. In the music education sector, the assistance of computer technology provides significant support. In literature [13], computer-assisted digital technology makes contributions to the development of students' thinking ability and innovation consciousness by means of converting the abstractness of music material into concrete and vivid perceptual material. Literature [14] designed a digital music curriculum system under the guidance of computer-aided instruction, realizing the digitization of audio, sonic visualizing and editing, recording the course of students' training and abstracting music theory into visualized images, which is helpful for improving the effectiveness of music learning. Literature [15] demonstrates that computer-assisted music flipping subject significantly enhances the level of students' music knowledge, skills, and performance. Literature [16] constructs a learning resource library of ethnic vocal music and personalized teaching system in the computer vision technology-assisted context, assists in classroom practice by means of image recognition algorithms, and makes great contribution to the digital and sustainable development of ethnic vocal music teaching. Literature [17] uses the rough-bipolar neutral set approach to assess the effect of multimedia teaching strategy in music education, such as gamification platform, virtual reality, etc., and concludes that computer-assisted tools make great contributions to bridging the gap between music theory learning and practice. Literature [18] shows that computer music technology-assisted education has greatly improved students' involvement, skill level, and classroom atmosphere, and the corresponding challenges involve differences in technology access and changes in teachers' roles. Literature [19] proves that as computer-assisted tools, digital games can effectively combine formal and informal music education and make great contributions to the enhancement of learning motivation and efficiency, while they are often faced with problems such as the availability of tools and costs, and in-depth study should be strengthened in the future.

In preschool music teaching, literature [20] explores the use of augmented reality and rapid response code technology-assisted early childhood music education interventions based on a gamified collaborative learning framework, and practice has shown that this tool can effectively stimulate children's interest and promote their cognitive, collaborative, and social development, demonstrating its potential for affective and cognitive development. Literature [21] developed a computer-assisted digital music teaching and learning material, which can effectively enhance the auditory and visual perceptual abilities of children with intellectual disabilities, especially in the instrument module, and also showed positive effects in sound recognition. Literature [22] has shown that computerized music teaching brings new opportunities for children's “listening to music” programs, which can effectively develop children's creativity and meet the requirements of modern pedagogy. Literature [23] explored the effects of digital technology in preschool music education, and found that the technology can enhance the development of children in the three dimensions of music knowledge, emotional engagement and creativity, which provides a basis for the development of relevant curricula and policies.

Based on basic music theory knowledge, this paper explains the two forms of data representation in computing, namely MIDI format and ABC format, and focuses on MIDI technology to carry out research on computer-assisted music teaching. Based on the original recurrent neural network model, adversarial generative network and variational autoencoder network are introduced to form a hybrid neural network (RNN-VAE-GAN) model to generate monophonic MIDI music, and the MIDI note features are extracted to optimize the MIDI auto-composed music model through WaveNet. The simulation dataset is constructed, and after verifying the performance of the model through simulation experiments, it is put into the actual music teaching classroom to analyze the students' music learning.

2 Teaching music theory based on MIDI technology

2.1 Representation of music data in computers

2.1.1 MIDI format

The beginning of MIDI will contain the complete track type, the total number of tracks, the length of the base note, and so on. In a MIDI document, a tick is the shortest unit of time it can recognize, and the number represented by the base length of a measure is the number represented by a quarter note in the MIDI document. This is because the number of beats (BPM) and the number of taps changes from phrase to phrase. If the number of beats in a minute is represented by BPM, and P is used to represent one beat, or P times, and t is used to represent the number of quarter-note ticks, then the absolute time of a tick is such that equation (1):

$$T_{tick} = \frac{60}{BPM} \div \frac{4t}{P} = \frac{15P}{t \times BPM} \sec ond \tag{1}$$

MIDI uses event messages to record and express the playing of MIDI instruments. In the case of a MIDI keyboard, a button press event will display the track number on which the event occurred, the number of ticks from the start of the entire recording (indicating time), the name of the event, the pitch, and the key velocity (indicating volume).

2.1.2 ABC format

Electronic sheet music in ABC format is a shortened form of sheet music, a format designed to mark music using plain text.ABC notation records pitches using letters such as CDEFGAB, and indicates information such as ascending and descending keys, tonal values, intonation, and ornamental tones by adding additional symbols to the notes.

2.2 Monophonic MIDI music generation based on hybrid neural networks

2.2.1 Music generation

(1) RNN network and Char-RNN network

Large-scale sequence data processing is not a good fit for RNN networks since they are part of binary neural networks, which have just one input and one output. Primitive recurrent neural networks (RNN), long short-term memory networks (LSTM), and gate-controlled recurrent units (GRU) are examples of recurrent neural networks that are often employed.

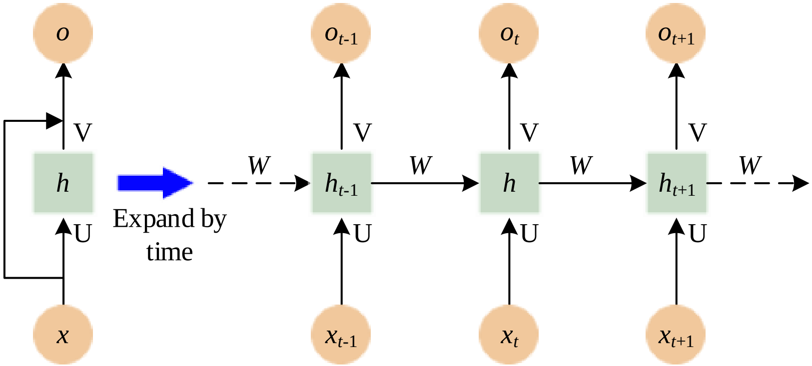

Figure 1: RNN model

Figure 1 shows the structure of an RNN network, where x_t is the input of the tth layer, h_t is the hidden state of the tth layer, which is jointly determined by the hidden state of the previous layer, h_{t-1}, and the input x_t of this layer, and o_t is the output of the tth layer. Throughout the cycle of the model consisting of multiple layers, the linear relationship parameters U, V, and W of the model are the same between each layer.

(2) LSTM network

LSTM network consists of 3 inputs and 2 outputs, Fig. 2 shows the construction of one of the nodes.In LSTM, in addition to the implicit layer state h, there exists an additional cell state C, which is used to record sequential information. The system employs three types of gates, namely “forget gate”, “input gate” and “output gate”, to realize the control of data transmission. Each gate structure controls the rate of information transfer with values in $[0,1]$ intervals. 0 means complete abandonment and 1 means complete preservation.

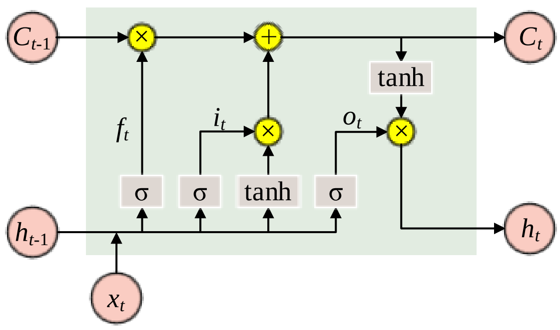

Figure 2: LSTM single-node structure

f_t is the forgetting gate, which indicates the proportion of C_{t-1} forgotten in the previous moment, and is calculated as shown in equation (2):

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right) \tag{2}$$

where W_f, U_f, and b_f are the model trainable parameters, and f_t is used as the output of the Sigmoid function, which takes values in the range $[0,1]$.

The structure of the input gate is shown in the middle of this unit, which consists of two parts: the input information $\tilde{C}_t$ and the transfer ratio of the input information i_t, both of which together represent the input information at that moment, and its calculation formula is shown in Eq. (3) and Eq. (4):

$$i_t = \sigma\left(W_i h_{t-1} + U_i x_t + b_i\right) \tag{3}$$

$$\tilde{C}_t = \tanh\left(W_C h_{t-1} + U_C x_t + b_C\right) \tag{4}$$

where W_i, U_i, b_i and W_C, U_C, b_C are model trainable parameters.

The forgetting gate and the input gate together determine the state of the cell at the moment t, as shown in Equation (5):

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{5}$$

The state of the hidden layer at the t moment is then determined by the output gate, as shown in Eqs. (6) and (7):

$$o_t = \sigma\left(W_o h_{t-1} + U_o x_t + b_o\right) \tag{6}$$

$$h_t = o_t \times \tanh\left(C_t\right) \tag{7}$$

W_o, U_o, and b_o are model trainable parameters.

As a whole, all RNN class neural networks suffer from the problem of gradient explosion and gradient vanishing, which leads to the problem that the generation quality of this type of composing model decreases dramatically as the length of the generated sequence increases.

(3) Adversarial Generative Networks

Adversarial Generative Networks (GANs) are one of the most promising non-supervised learning algorithms currently available for complex distributions.

When the discriminator determines that the input sample is a true sample, the discriminator will output a flag bit of 1, and vice versa if the input sample is judged to be a generated sample. In the process, between the generator and the discriminator for a game of adversarial behavior.

The loss function of the GAN model is shown in equation (8):

$$\begin{aligned}\min_G \max_D V\left(D,G\right) = {} & E_{x\sim p_{data}(x)}\left[\log D\left(x\right)\right] \\ & + E_{z\sim p_z(z)}\left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right]\end{aligned} \tag{8}$$

(4) Variational Auto-Encoder Networks

Variable Auto-Encoder Networks (VAE), one of the more popular architectures in music generation modeling nowadays, evolved from the Auto-Encoder (AE):

The loss function of a VAE consists of two components, the mean variance and the KL dispersion (KLD) between x_t' and x. The KLD measures the similarity of the two distributions, whereas in the VAE model the KLD is employed to measure the implied and normal distribution differences.The formula for the KLD is given in Equation (9):

$$KL\left(P \| Q\right) = \int P(x) \log \frac{P(x)}{Q(x)} dx \tag{9}$$

where $P(x)$, $Q(x)$ represent two probability distributions on the random variable X, $x \in X$.

2.2.2 Music Generation Modeling with Hybrid Neural Networks

(1) Processing of the training set

The main contents of the MIDI file are: pitch, onset and time of the notes. On this basis, the syllables in the text were analyzed and filtered to obtain the corresponding training results. The obtained parameters were expressed in a two-dimensional matrix whose two dimensions represent a pitch and a moment. Here, in MIDI format, the tick (the shortest moment in a midi document) is quantified, with a minimum of sixteen bytes.

(2) Encoder structure

The MIDI music messages in the training set are first stored in a four-dimensional matrix, where the value of each element of the matrix is the note, pitch, onset time, and tense value of the note for each time step. The n -bar music sequence is represented by $X=\{x_0,\cdots,x_{t-1},x_t,\cdots,x_n\}$, where x_{t-1} and x_t are two consecutive pieces of information.

Next, the music information of the current moment is encoded as a one-dimensional latent vector by the encoder, together with the music information input at the previous point in time. This music information consists of information such as the note, pitch, start time, and the time value of the note. Fig. 3 shows the construction of the encoder using a single layer of bi-directional GRU units.

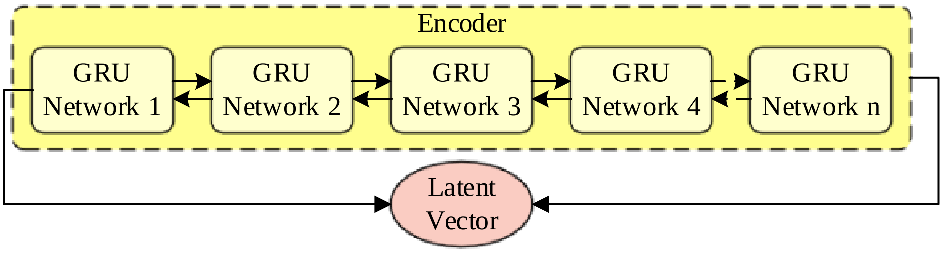

Figure 3: Encoder structure

Gated Recurrent Unit Recurrent Neural Network GRU is a special RNN network structure with the same network update method as RNN, although GRU is designed to solve the long dependency problem in RNN networks.

Figure 4 shows the GRU unit structure, compared to RNN, GRU adds two gate control units, an update gate, and a reset gate.

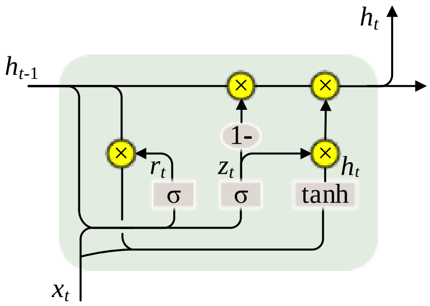

Figure 4: GRU unit structure

The constituent functions are as follows:

$$z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \tag{10}$$

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \tag{11}$$

$$y_t = \sigma\left(W_o \cdot h_t\right) \tag{12}$$

$$\tilde{h}_t = \tanh\left(W_h \cdot \left[r_t * h_{t-1}, x_t\right]\right) \tag{13}$$

$$h_t = \left(1 - z_t\right) * h_{t-1} + z_t * \tilde{h}_t \tag{14}$$

where: z_t denotes the update gate at moment t, r_t denotes the reset gate at moment t, $\tilde{h}_t$ denotes the candidate activation state at moment t, h_t denotes the activation state at moment t, h_{t-1} denotes the hidden layer state at moment $t-1$, x_t denotes the input at moment t, σ denotes the activation function, and W_r, W_z, W_h, and W_o are all weight parameters to be learned. The reset gate r is based on the information obtained by the candidate state from the previous information, while the update gate z is based on the previous information as well as the requirement that the current state forget itself and the new knowledge it possesses. The update gate regulates how much data is sent from the previous time step; the greater the update gate's value, the more data is sent. The reset gate regulates how much data is written to the current time step from the previous time step; the greater the reset gate's value, the more data is written to the current time step.

(3) Generator (decoder) structure

Through the generator, the low-dimensional music feature information is extracted from the latent vectors and decoded into new music information. Since the decoder is built as a two-layer GRU, the first layer GRU extracts the low-dimensional music feature information from the latent vectors and maps it into U local vectors through the network of this layer, and then uses these U local vectors to initialize the network of the second layer GRU, and with such a design, the decoder is able to obtain the long term context dependency through these vectors. The decoder samples latent vectors from the latent vector distribution $P(z)$.

Finally, x is sampled from the distribution $P\left(x; g\left(z\right)\right) = P\left(x|z\right)$, x denotes the original data sample, $P(x)$ is the distribution of the original data sample, z is the latent vector, and $g(z)$ denotes the latent variable distribution.

(4) Discriminator Structure

In the training process, the ability of music generation module and music evaluation module is not always equal, it often appears that the music generation module or music evaluation module is trained to be so powerful that the gradient of the other side disappears, which is one of the main reasons for the unstable quality of GAN network compositions. Therefore, a freezen mechanism is added to improve this problem: when a party is too powerful to make the training gradient disappear, the overly powerful party will be “frozen”. Its objective function is:

$$\min_G V(D,G) = E_{x\sim p_{data}(x)}\left[D(x)\right] - E_{x\sim p_g}\left[D(x)\right] + \lambda E_{pg}\left[\left\|\nabla_x D(x)\right\|^2\right] \quad (15)$$
$$\min_G V(D,G) = E_{z\sim p_z}\left[D\left(G(z)\right)\right]$$

where D denotes the music evaluation function, G denotes the music generation function, x denotes the real data input, $E_{x\sim p_{data}(x)}$ denotes the sampling of x from the distribution pdata, data denotes the real data, $pdata(x)$ denotes the distribution of the real data data, z denotes the noisy data, p_z the distribution obeyed by the noisy data, and p_g is the distribution obeyed by the generated data. $D(x)$ denotes the expectation of x when x obeys the distribution of pdata, and the output is a value with a maximum value of 1 and a minimum value of 0. λ is the penalty term $\lambda E_{pg}\left[\left\|\nabla_x D(x)\right\|^2\right]$ parameter.

2.3 MIDI music generation optimization

2.3.1 Automatic compositional note feature extraction

(1) MIDI dataset analysis

The Enya MIDI Music Data Set is a collection of digital music files of multi-instrument ensembles that is utilized in automated composition research. The great majority of the musical pieces in the data set are composed of multiple instruments, and each song's various measures are either performed by individual instruments alone or by multiple instruments in the form of an orchestra, which can occasionally be graceful, passionate, serene, and lively. Additionally, a 4/4 time beat is adopted by most of the songs in the data set.After learning, it is possible to incorporate each instrument's unique qualities into the process of making multi-instrument orchestras. As a result, using this data set has the following advantages: first, it contains a variety of musical instrument types, which makes it easier to study multiple instruments and multi-instrument orchestras; second, most of the songs in the data set have a 4/4 time rhythm, which makes it easier to combine instruments when building orchestras; and third, it uses a MIDI file format, which makes it easier to analyze and extract features.

(2) MIDI feature extraction

Since the monophony and chords in the file are both the input of our automatic composition algorithm and the output of music production, extracting the monophony and chords from MIDI files is the main challenge when it comes to feature extraction. Ultimately, all of the pitch and chord predictions will be serially arranged into a track object to generate the performance score of a single instrument.

2.3.2 Automatic compositional modeling

We obtained both monophonic and harmonic sequences for every musical instrument in the database after the feature extraction method of the MIDI files. The principle underlying the automated composition used in this work is quite similar to typing, where the next word is suggested by the input method. For this project, an array of monophonic and chordal sequences of length 32 is chosen as the input of the model, and the next monophonic or chordal sequence that is about to appear is predicted, and the result of the prediction is taken as the output, and then the output is taken as a part of the array of pitch sequences as the input to continue the prediction and so on, and then a sequence of predictions is obtained in a circular manner, which is the pitch data of the result of composing.

The basic principle of WaveNet is to predict the value of the first t point based on the first $t-1$ points of the sequence, therefore, when we know the sequence of notes in a piece of music, we can use the model to predict the subsequent possible notes, the basic formula of WaveNet prediction is shown in Eqn. 16:

$$p(x)=\prod_{t=1}^{T} p\left(x_t \mid x_1, x_2, \cdots, x_{t-1}\right) \tag{16}$$

WaveNet is a convolution-based model having a multi-layer convolution neural network architecture. In the WaveNet architecture, each of the convolution layers conducts a convolution operation on the output of the preceding layer, and the more the size of the kernel of the operation being conducted, the more will be the perceptual range of the layer, making it have more time domain perception capability. Once the output layer receives the final result of the operation done by the convolutional layer, it gives out that result and makes use of it as an input to conduct operations again in an endless cycle is shown in Figure 5.

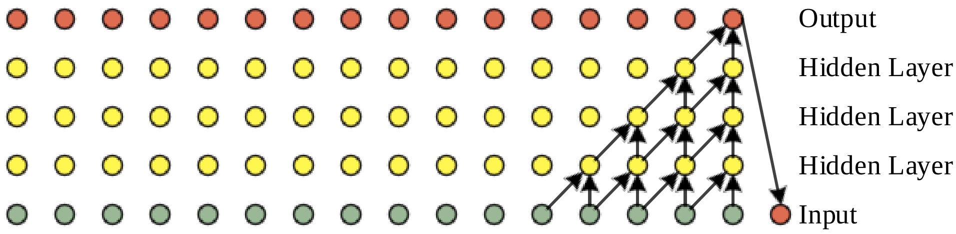

Figure 5: Schematic diagram of the WaveNet principle

Based on our thoughts and considerations, we conclude that the use of only one activation function in the model, whose richness needs to be enhanced, along with the fact that the fully connected layers in the model are very few, leading to large prediction errors, may be the primary cause of the delayed decrease of the loss value in the training dataset. Based on the research on the proposed model and the information from consulting, we learn that there are numerous reasons behind this. Therefore, we propose the following optimizations: 1) Increase the number of fully connected layers, in order to reduce the error of model outputs.

1) Enhance the number of fully connected layers to reduce the error of the model output results.

2) Adjust the number of neurons in each connection layer to improve the accuracy of neuron output.

3) The model is better equipped to replicate the playing characteristics of each specific musical instrument thanks to the use of several activation functions. The convergence effect of the model will not be optimal if there are too many fully connected layers and neurons, despite the fact that there are plenty of fully connected layers and a large number of neurons, according to the aforementioned ideas and numerous comparative experimental trials. When compared to the initial model throughout the training process, certain models that have been altered in line with specific musical instruments exhibit greater rates of loss. Consequently, the number of neurons and the number of completely linked layers should be chosen sensibly. In the meanwhile, it has been noted that before the loss rate stays almost constant, a specific number of training cycles must be completed. Therefore, in order to decrease training time, fewer training rounds should be used. In conclusion, the following enhancements are made:

1) Assign three levels to the completely linked layer.

2) Set the number of single tones and chord types, as well as the number of neurons, to 3000 and 1024, respectively.

3) Assign the tanh activation function to the newly added completely linked layer.

4) Reduce the number of training cycles from 300 to 150, using 30 as a checkpoint.

3 Experimental results and analysis of MIDI technology

3.1 Experimental conditions

3.1.1 Experimental data set

The 391 transposed and filtered MIDI files of the piano from chapter 2 served as the training set for this experiment. One of the two guidelines for choosing the music is to pick piano pieces with a 4/4 beat. Music compositions with a major pentatonic or minor pentatonic scale percentage rate of 80% or above for their main theme were selected. Since the field of music generation is still in its infancy, no testing datasets are used.

3.1.2 Automatic Composition Experiment

Using the RNN-VAE-GAN piano transcription model, piano music data in mp3 format collected from major music platforms are converted to MIDI format. Where a2m.pth is the weight file of the transcription model, the inference calculation related to the transcription is done by the GPU.

3.2 Analysis of Composition Results

3.2.1 Number of notes for generating music samples

The first comparative experiment was conducted in the objective comparison of producing music. Up to 200 bits of music were created by four models, totaling 50 compositions. Sample 1 was the music produced by the C-RNN-GAN model; Sample 2 was the music produced by the Music-VAE model; Sample 3 was the music produced by the RNN-VAE-GAN model discussed in this paper; and Sample 4 was the music produced by the Pop music transformer model. In contrast to the measurements used in this research, some examples of potential metrics for objectively assessing the quality of automated compositions are given. The number of notes measure, which is given as the number of notes in 16 bars, serves as an illustration of one such measure. In this context, Figure 6 shows the number of notes in 16 bars for each of the 50 compositions of the four track samples.

The average notes number in the songs generated by the C-RNN-GAN automatic music composition model is at least 158.39 and comparatively low, according to the Notes Number data, indicating that music emotion is expressed slowly and softly. The greatest note number in music production, 199.19, is comparatively high when compared to the Music VAE auto-composition model, indicating that music is created with greater intensity. Furthermore, the RNN-VAE-GAN and Pop music Transformer models produce music with a greater standard deviation and, in some situations, a higher or lower note number than the other two RNN-based models. In this instance, the RNN-VAE-GAN model has 184.24 notes, which indicates that the music generated by these models is rich in emotion expression, distinct rhythm, and emotion production.

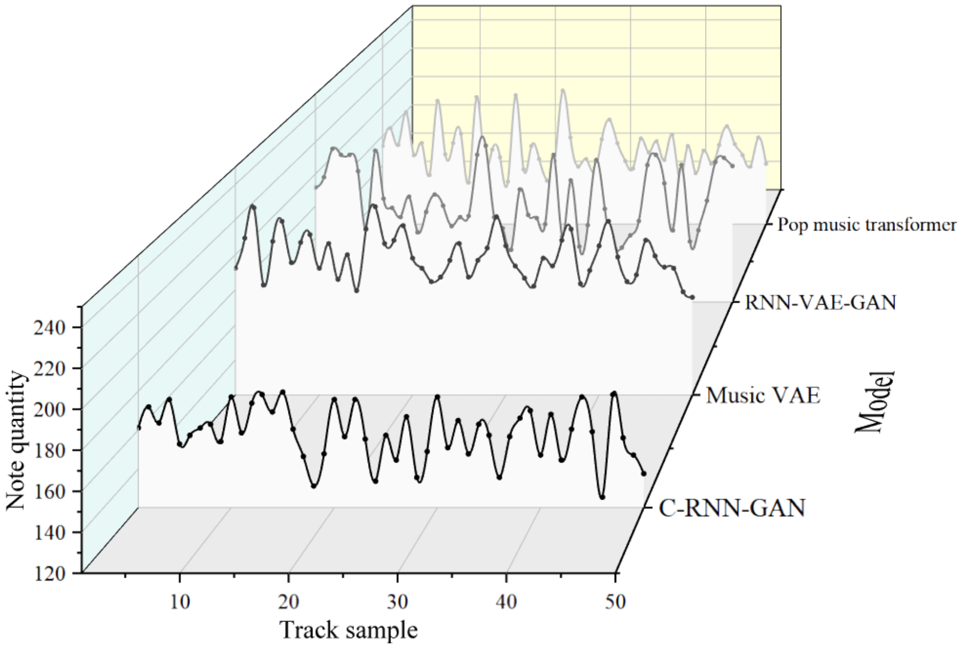

Figure 6: The model generates the number of notes of the music sample

3.2.2 Pitch

The notes in the pentatonic score become points on the coordinates when seen as coordinates, and all of these points are eventually connected to form a melodic line of sounds. Both the note intervals and the musical movement from one place to another are visible in the melodic line. Plotting the pitches included in the primary melody line of the produced music, as seen in Figure 7 below, reveals that the musical pitches follow the note trajectory to a very high degree of 98%.

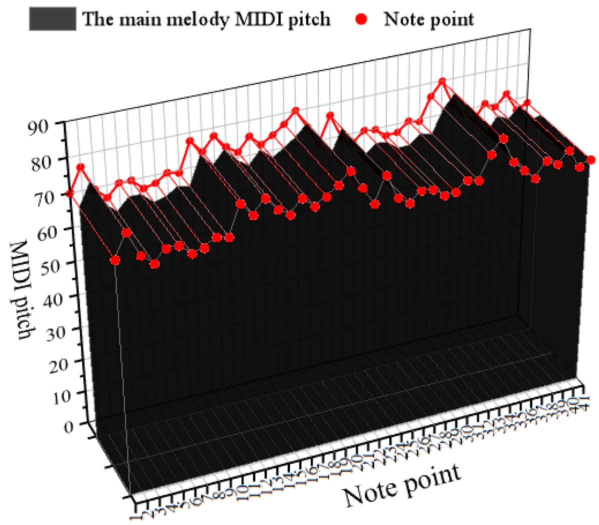

Figure 7: Generate the pitch intervals of the main melody of the music

The next objective evaluation is the distribution of note pitches, and the histogram of pitch distribution for the four track samples is shown in Figure 8. The histogram of pitch distribution from above visualizes two characteristics of the generated music, one is the tonality of the generated music, and the other is the range of the generated music. The samples of the tracks generated by the four models have a large number of pitches concentrated in the mid-region of the piano in the regions of C, D, E, G, and A. The range of the C-RNN-GAN is from E2-#F6, the range of the Music VAE is from C2-E6, the range of the RNN-VAE-GAN is G1-B6, and the range of the Pop music transformer is G1-B5. This interval relationship corresponds to the pitch range of the music transformer in the region of G1-#F5. This interval relationship corresponds to the C major pentatonic scale with C as the dominant or the A minor pentatonic scale with A as the dominant. The automatic composition model based on the C-RNN-GAN network is more obvious in this performance, with a narrower range and a pitch ratio of 2.014, which indirectly indicates the tonal homogeneity of its generated music. In terms of tonal range, the RNN-VAE-GAN auto-composition model has the widest range in generating music, with a pitch percentage of 1.57%, which indicates its use of richer tones.

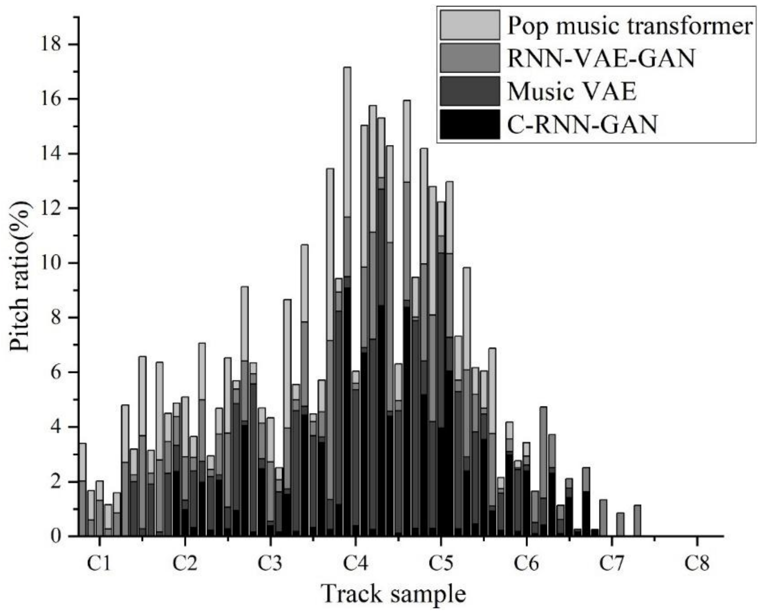

Figure 8: The model generates the pitch of the music sample

3.2.3 Objective evaluation indicators

Regarding the music in this paper that is being examined, the final objective evaluation experiments were conducted by measuring the main melody's pentatonic scale rate. The statistics of the main melody's major and minor pentatonic scale rates in the sample pieces of music were gathered independently. Regarding the models, they were all trained using music data sets where the primary theme's pentatonic scale rate was more than 80%. Therefore, we will use the pentatonic scale rate of 80% as a reference for our experiment, and we will filter out the number of musical compositions whose main melody's major or minor pentatonic scale rate surpasses 80% while also accounting for their maximum values. The final results are shown in Table 1:

The auto-composition model of RNN-VAE-GAN outperforms other models in music generation based on the pentatonic scale rate experiment results. It can generate track samples of three different types of music with major pentatonic scale rates greater than 90%, with the highest values being 91.24% for the major pentatonic scale rate and 86.26% for the minor pentatonic scale rate, which are nearly equal to those of human work pieces.

Table 1: Objective evaluation index of the main melody pentatonic scale

Track sample number	The number of changes in the number of pentatonic scales is greater than 80%	The number of features of the small tone is greater than 80%	The maximum value of the pentatonic scale in major	The highest pentatonic scale ratio in minor Large value
1	26	5	86.32%	83.85%
2	11	4	82.32%	81.25%
3	26	13	91.24%	86.26%
4	22	8	85.36%	84.42%

Based on the scores of the music of each compositional approach is summarized and finally shows the average score, as well as the highest and lowest scores, the different compositional approaches and the resultant manual evaluation scores are shown in Table 2. As far as the manual subjective evaluation results are concerned, it shows that among the automatic composition models, RNN-VAE-GAN has the highest average score of 3.79, while Pop music transformer comes second with 3.65. 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.

The results of several experiments in this section finally show that the RNN-VAE-GAN automatic composition model designed based on RNN networks performs the best in terms of generating music quality (mainly in the five aspects of the number of notes, the average interval interval, the pitch distribution, the rate of pentatonic scale, and the auditory evaluation), and it can be used for the teaching of pre-school education in music theory.

Table 2: Artificial evaluation score

Scoring grade	Description	Composition method	Average score	The highest score	The lowest score
5 points (excellent)	Excellent quality	Artificial composition	4.36	5	3
4 points (Good)	Good quality	C-RNN-GAN	3.32	4	2
3 points (medium)	The quality is average	Music VAE	3.08	3	2
2 points (short)	Poor quality	RNN-VAE-GAN	3.79	5	3
1 point (poor)	Extremely poor quality	Pop music transformer	3.65	4	3

3.3 Instructional design for preschool music theory

3.3.1 Instructional design

This study designs a computer-assisted teaching model for pre-school music teaching by combining the computer-assisted teaching models constructed by previous researchers. In this model, it is mainly divided into three main parts: before class, during class and after class. Since teaching activities are composed of “students' learning and teachers' teaching”, this study will elaborate on the before class, during class and after class parts of the computer-assisted teaching model from the perspective of teachers.

It should be noted that in the computer-assisted preschool music teaching model, the “restricted access” condition is a prerequisite for class formation. According to the class division of compulsory education, the class size of each school is about 50-70 students, and the learning level of the students in the class is basically the same, so the researcher believes that the class can be taken as a unit, and “S” is interpreted as “small-sized classroom teaching”, which means that all the students can have access to the computer-assisted pre-school music teaching model. All access.

(1) Pre-lesson

In the pre-class session, teachers need to introduce appropriate online learning resources, make a list of students' learning tasks, design relevant questions and online quizzes according to the learning situation, textbook content and music teaching objectives, and record and summarize the quality of students' online learning and the questions they ask in order to prepare for the preparation of the class lesson plan.

(2) In-class

The main way to carry out different exercises and finish bilateral teaching tasks is in the classroom. While conducting the process of internalizing knowledge through group communication and activities, teachers should conduct classes based on the quality and state of their online education. Additionally, during the interaction process between teachers and students, the former will receive feedback from the latter regarding the unsolved problems.

(3) After class

After completing the classroom offline teaching, the teaching session is not all over. According to the Ebbinghaus forgetting curve, students always forget the newly learned knowledge faster in the initial stage, so they need to consolidate and improve their knowledge in the after-class session.

3.3.2 Instructional evaluation design

The teaching evaluation in this study will be based on the characteristics of the computer-assisted music teaching mode and the preschool music theory subject, and will be designed from the three aspects of evaluation content, evaluation subject, and evaluation method according to the new Art Standards in 2022.

Two parallel classes were selected as the experimental class and the control class to carry out a six-week comparative experiment, combining the more typical examples in music textbooks to elaborate the specific implementation process based on the computer-assisted teaching model in the preschool music classroom.

Pre-test and post-test techniques were used in the experiment design, and two parallel classes were selected as the experimental group and control group for the comparative teaching procedure. Before the beginning of the experiment, teachers need to conduct a music test and complete a questionnaire on music learning for students in the experimental class and the control class. The music test is used to test whether the music level of students in the two classes is the same and whether they meet the conditions of the experimental subjects, and to facilitate comparison after the experiment. The questionnaire was used to find out the students' situation in the four dimensions of music learning interest, music learning habits, music core literacy, and classroom satisfaction. At the start of the experiment, the instructor used computer-assisted instruction in the experimental class whereas the control class used traditional instruction. Following the experiment, the same questionnaires were administered to both the experimental and control groups in order to ascertain whether there had been any notable changes in the four dimensions of students' interests in learning music. Additionally, a test was conducted to compare and analyze the differences between the two groups' musical proficiency using the music test paper in order to ascertain whether the experimental group's performance had improved under the influence of the computer-assisted teaching mode and to assess its efficacy in teaching music to preschoolers.

3.4 Analysis of the Effectiveness of MIDI-based Preschool Music Theory Instruction

3.4.1 Comparative analysis of students' academic performance

The sample students in the two classes were assessed and tested for musical theories, musical professional skills, musical learning interest, musical learning habits, musical essential attributes, and class satisfaction after six weeks of teaching practice. The results were then subjected to the independent sample t-test and the paired-sample t-test for the pre-experimental and post-experimental data of the two student classes.

At the end of the teaching practice, the two groups of students took separate examinations under the conditions of the autonomous section. These assessments included professional music abilities (50 points) and basic music theory knowledge (50 points). Following this stage, the test phase data for the experimental group and the control group were acquired, respectively. To get the same assessment score for the test paper, sample pairing was done after the test data was processed. The analytical findings of the two sections' independent sample t-tests, as well as the overall scores of the experimental and control groups, are shown in Table 3. The test findings showed that the control class had an average score of 76.03±2.42 with a significance level of 0.000, whereas the experimental class had an average score of 79.06±4.36. The score results indicate a statistically significant difference between the experimental class and the control class with respect to the threshold level of 0.05, and the experimental class's professional musical skills and music theory score are statistically significantly higher than those of the control class. Thus, we may conclude that the experimental class's use of computer-assisted instruction enhanced students' performance in music theory and abilities.

Table 3: Independent sample t-test

Indicator heart	Class		T	P
	Experimental Class	Control class		
Basic knowledge of music theory	39.48±3.42	38.48±2.54	3.048	0.0024**
Music professional skills	39.58±3.12	37.55±2.33	2.536	0.013*
Total score	79.06±4.36	76.03±2.42	5.415	0.000***

3.4.2 Comparative analysis of students' music learning in the two classes before the experiment

The four dimensions of the experimental and control groups' pre-test scores are displayed in Table 4 using the t-test for independent samples. The P value for the two groups' enthusiasm in studying music is 0.948, the P value for their learning habits is 0.452, the P value for music core literacy is 0.963, and the P value for classroom satisfaction is 0.785, according to the data in the table following the questionnaire analysis. The P values of the four dimensions are all larger than 0.05, using 0.05 as a reference standard for the significance threshold. This indicates that the pre-test scores of the experimental group and the control group do not differ significantly.

Table 4: Test of independent sample t of pre-measured data

Indicator	Class		T	P
	Experimental Class	Control class		
Classroom satisfaction	3.01±0.24	2.94±0.22	0.348	0.785
Core musical literacy	2.53±0.36	2.54±0.23	0.018	0.963
Music learning habits	2.34±0.29	2.35±0.25	0.536	0.452
Interest in learning music	2.64±0.24	2.64±0.28	0.023	0.948

3.4.3 Comparison of students' music learning in the two classes before and after the experiment

(1) Control class

The paired sample t-test used to compare the pre-test and post-test findings in each of the four dimensions examined in the control class is shown in Table 5. The numbers from the control group on each of the four dimensions—musical interest, habits, literacy, and classroom satisfaction—vary somewhat, but they do not differ significantly. The p-value of the control class is 0.432 for students' learning interest, 0.000 for music learning habits, 0.035 for music core literacy, and 0.812 for classroom satisfaction, and with the value of 0.05 as a reference for the significance value, the post-test level of the control class is higher than the pre-test level in the two dimensions of music core literacy and music learning habits, and the post-test level in the two dimensions of music learning interest and classroom satisfaction dimensions are not significantly different, indicating the lack of traditional teaching methods in classroom satisfaction and music learning interest.

Table 5: Comparison of the results of the comparison

Indicators	Class			T	P
	Pre-test	Post-test	Post-test - pre-test		
Classroom satisfaction	2.98±0.23	3.01±0.38	0.03	0.245	0.812
Core musical literacy	2.54±0.34	2.64±0.48	0.1	2.096	0.035*
Music learning habits	2.38±0.23	2.63±0.34	0.25	5.015	0.000***
Interest in learning music	2.65±0.27	2.67±0.12	0.02	0.896	0.432

(2) Experimental class

Table 6 compares the experimental class's pre-test and post-test outcomes. Regarding the four aspects of music learning interest, music learning habits, music core literacy, and classroom pleasure, there are clear shifts in the values of the students in the experimental class. The P-value of the students in the experimental class for classroom happiness and music learning habits is 0.000, indicating that all improvements are significant. The music core literacy P-value is 0.015, but the music learning interest P-value is 0.002. The experimental class's post-test level clearly exceeded the pre-test level, using the critical value of 0.05 as the reference for the significance level. This suggests that all four dimensions in the experimental class have clearly improved after the training experiment. As a result, the use of computer-assisted instruction has enhanced music core literacy and classroom satisfaction while also positively stimulating students' enthusiasm in studying music and helping them build strong music learning habits.

Table 6: Comparison of test results before and after the experiment in the experimental class

Indicators	Class			T	P
	Pre-test	Post-test	Post-test - pre-test		
Classroom satisfaction	3.01±0.24	3.48±0.34	0.47	8.265	0.000***
Core musical literacy	2.48±0.26	2.64±0.35	0.16	2.482	0.015*
Music learning habits	2.48±0.25	3.25±0.33	0.77	16.254	0.000***
Interest in learning music	2.67±0.36	2.86±0.46	0.19	3.059	0.002***

3.4.4 Comparative analysis of students' music learning in the two classes after the experiment

The comparative analysis of the students' music learning environments in the two classes during the experiment is shown in Table 7. The results indicate that the P-value of the two variables is less than 0.001, while the P-value of the students' classroom satisfaction and music learning habits in the two classes is 0.000. This indicates that there is a very big difference. On the other hand, the P-values for music learning interest and music core literacy are 0.001 and 0.005, respectively. In contrast, a significant difference is shown by the P-value of 0.05, which is less than 0.05. This is due to the experimental class's superior performance employing the computer-assisted teaching paradigm.

Table 7: Comparative analysis of music learning in the experiment

Indicators	Class		T	P
	Experimental Class	Control class		
Classroom satisfaction	3.01±0.24	3.01±0.38	7.145	0.000***
Core musical literacy	2.48±0.26	2.64±0.48	2.933	0.005**
Music learning habits	2.48±0.25	2.63±0.34	10.236	0.000***
Interest in learning music	2.67±0.36	2.67±0.12	3.348	0.001**

4 Conclusion

In this paper, we explore the knowledge of music music theory from four aspects of melody, rhythm, pitch combination and melodic movement, propose two formats of MIDI and ABC for music data representation in computers, combine monophonic MIDI music with hybrid neural networks to automatically generate music melodies, and propose optimization schemes for MIDI music generation techniques. Construct experimental dataset and implement automatic composition experiments using RNN-VAE-GAN piano transcription model.

The RNN-VAE-GAN architecture suggested in this study generates 184.24 total notes according to objective standards. The RNN-VAE-GAN model's total notes have a bigger variance than those produced by other RNN architectures. This indicates that the music produced has a high degree of emotional expressiveness, making it more suitable for producing a variety of feelings.

Designing a preschool music theory teaching experiment to analyze the computer-assisted teaching strategy based on comparing the music learning of students in the experimental class and the control class, the p-value of music learning habits and classroom satisfaction of students in the two classes is 0.000, and the p-value of the two dimensions is less than 0.001, which is a very obvious difference, and the experimental class that implements the computer-assisted teaching model has higher progress.

About the Author

Peijun Xie, born in Ankang, Shaanxi, China in November 1976, graduated from Xi'an Conservatory of Music with a Bachelor's degree. She is currently an associate professor at Ankang Vocational and Technical College, mainly engaged in research on music in preschool education.

References

[1] Baxodirovna, A. Z. (2022). International trends in the development of the preschool education system. INTERNATIONAL JOURNAL OF SOCIAL SCIENCE & INTERDISCIPLINARY RESEARCH ISSN: 2277-3630 Impact factor: 8.036, 11(11), 458-461.

[2] Guerrero, S., Nunez, M., & Corbacho, C. (2023). Shaping executive function in pre-school: The role of early educational practice. Cognitive Development, 67, 101344.

[3] Foti, P., & Sidiropoulou, M. (2019). THE VALUE OF PRE-SCHOOL AND EDUCATION-HISTORICAL REVIEW AND CONTEMPORARY TRENDS IN PRE-SCHOOL EDUCATION. The Cyprus Journal of Sciences, 17, 3-15.

[4] Penney, S. C., Young, G., Butler, E., Maich, K., & Philpott, D. F. (2019). The role of quality ECE in facilitating mental health and well-being in children. Exceptionality Education International, 29(3), 57-76.

[5] Sušić, B. B. (2019). PRESCHOOL EDUCATION STUDENTS'ATTITUDES ABOUT THE POSSIBLE IMPACT OF MUSIC ON CHILDREN'S SPEECH DEVELOPMENT. International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE), 7(1), 73-84.

[6] Kayili, G., & Kuşcu, Ö. (2021). The effects of Orff-based attention-enhancing music education programme on impulsive preschool children's cognitive tempo. In Music in the Lives of Young Children (pp. 323-334). Routledge.

[7] Borawska, I. (2021). Music, visual arts, literature as educational means of holistic influence on a preschool child: stimulating children's development through art (Polish experience). Zhytomyr Ivan Franko state university journal. Pedagogical sciences, (2 (105)), 46-60.

[8] Jiang, F. (2023). Research on the role of preschool music education on the quality cultivation of young children. The Educational Review, USA, 7(1), 10-14.

[9] Boland, N. (2025). Music as expression of children's spirituality: the case of Steiner early childhood education. International Journal of Children's Spirituality, 30(1-2), 96-112.

[10] Don, G., Garvey, C., & Sadeghpour, M. (2023). Theory and practice: Signature pedagogies in music theory and performance. In Exploring Signature Pedagogies (pp. 81-98). Routledge.

[11] Harrison, D. (2020). Pedagogy into practice: teaching music theory in the twenty-first century. Journal of Music Theory, 64(2), 297-308.

[12] Maba, A. (2020). Computer-aided music education and musical creativity. Journal of Human Sciences, 17(3), 822-830.

[13] Yuan, Y. (2021). Design and Realization of Computer Aided Music Teaching System Based on Interactive Mode. Computer-Aided Design & Applications, 18.

[14] Pei, Z., & Wang, Y. (2022). Analysis of computer aided teaching management system for music appreciation course based on network resources. Computer-Aided Design & Applications, 19.

[15] Qisen, Z., Jamaludin, K. A., Nasri, N. M., & Jinglong, L. (2024). Computer-assisted education and application in the music flipped classroom: A review of the literature. Edelweiss Applied Science and Technology, 8(4), 1297-1308.

[16] Jia, H. (2023). Design and implementation of personalized teaching system for ethnic vocal music learning resources based on computer vision. International Journal of Educational Innovation and Science, 4(1).

[17] Wang, W. (2025). Rough-Bipolar Neutrosophic Set for Music Teaching Based on Multimedia Computer-Assisted Technology: A Study on Instructional Quality and Student Engagement. Neutrosophic Sets and Systems, 90, 868-877.

[18] Sularso, S., Yu, Q., & Pranolo, A. (2024). Advancing computer science in education: Integrating digital music technology into elementary school music programs. In E3S Web of Conferences (Vol. 501, p. 01019). EDP Sciences.

[19] Weatherly, K. I. C. H., Wright, B. A., & Lee, E. Y. (2024). Digital game–based learning in music education: A systematic review between 2011 and 2023. Research Studies in Music Education, 1321103X241270819.

[20] Preka, G., & Rangoussi, M. (2019, May). Augmented Reality and QR Codes for Teaching Music to Preschoolers and Kindergarteners: Educational Intervention and Evaluation. In CSEDU (1) (pp. 113-123).

[21] Sağer, T., & Şakalar, G. Y. (2018). An investigation on a computer-assisted digital musical teaching material for the special education of educable mentally retarded children in Turkey. Turkish Online Journal of Design Art and Communication, 8(3), 583-594.

[22] Gorbunova, I. B., & Kiseleva, Y. N. (2020). Music Computer in Teaching the" Listening to Music" Course. Journal of Educational Psychology-Propositos y Representaciones, 8(2), e522.

[23] Wang, L. (2025). Children's Enlightenment Music Education Based on Digital Technology. Journal of Computer Assisted Learning, 41(2), e70022.