Design and Implementation of Multimedia Teaching Course for Piano Enlightenment Oriented to Aesthetic Ability Development

Xuan Liu, Jinggangshan University, China*

ABSTRACT

Piano education for children is based on the child’s psychological characteristics and the artistic characteristics of musical works. Piano initiation education for young children can enhance perception, aesthetic ability, and creative skills; promote the overall development of children; and is also a crucial channel to promote young children to develop and transfer their own musical experience. This paper analyzes the current problem of cultivating children’s musical aesthetic ability in piano initiation education and proposes the use of multimedia and artificial intelligence technology to compare the repertoire played by piano learners with the original repertoire, combined with aesthetic evaluation technology to analyze and give guidance for piano initiation practice, so that piano learners can be more targeted. The program is designed to solve the problems in the piano practice process, to enhance the musical perception and aesthetic feeling of children, and to improve their aesthetic ability.

KEYWORDS

Aesthetic Development, Artificial Intelligence Technology, Beginning Piano Instruction, Multimedia Teaching

INTRODUCTION

With the strengthening of China’s economy and the progress of social development, people are paying more attention to artistic cultivation, and many parents of children are improving their children’s humanistic temperament by cultivating their piano performance ability, and piano education for young children is becoming more and more important. The current demand for piano education is relatively large, and the market for piano education is gradually expanding, so there are increasing demands for the quality of teaching. Piano education not only improves the child’s musical aesthetic ability, but also makes the child’s education more scientific and reasonable (Qureshi et al., 2019; Iwendi et al., 2019).

For young children, piano education can be tailored to meet their individual musical needs, allowing music education to reach all children. There are many differences among young children, and they also have different musical needs and music-related learning methods. Because of many
reasons such as heredity and living environment, each child’s specific development level is uneven. The requirements for music learning in a group setting are relatively uniform, whereas piano education can consider the individual differences of children to the greatest extent possible, so that children’s musical skills and levels can be enhanced as much as possible. The extensive development of piano education allows children to have a better understanding of music and to satisfy their individual differences, which helps to develop their aesthetic ability for music and, ultimately, to improve the effectiveness and quality of teaching (Siddiqui et al., 2017).

In the early childhood stage, when people are in a period of rapid physical and psychological growth, music education can promote the intellectual development, emotional cultivation, and artistic cultivation of young children (Lin et al., 2020). Music, as an auditory art, contains different styles, weaves, timbres, and rhythms, and requires young children to have certain aesthetic skills in order to appreciate the charm of music as a non-semantic language. Therefore, it is necessary to strengthen the development of young children’s musical aesthetic skills to ensure that they develop in many ways. With the rapid progress of multimedia and artificial intelligence technology in recent years, it has shown a very high development potential in several fields (Antoro et al., 2019). In the relatively complex field in chess, AlphaGo developed by Google successfully defeated the world champion Lee Sedol. In the field of speech recognition, more mature software already exists on the market, such as many voices assistant tools. Piano initiation education is no exception.

This paper analyzes the problem of cultivating children’s musical aesthetic ability in the current piano enlightenment education, and puts forward the use of multimedia and artificial intelligence technology to compare the music played by piano learners with the original music, and combine the aesthetic evaluation technology to analyze and guide the piano enlightenment practice, so as to make piano learners more targeted. This study aims to solve problems in piano practice, enhance children’s musical perception and aesthetic sense, and improve their aesthetic ability.

RELATED WORK

Research on the Piano Enlightenment Education for Children

In piano education, children need to learn how to read music with their eyes, play with their hands, identify sounds with their ears, pedal with their feet, and think with their brains, all of which not only promote intellectual development, but also enable children to use their visual perception to develop comprehensive visual thinking activities, thus strengthening their musical awareness (Raja et al., 2020). It has been proven that the age of 4 to 6 is the best listening period, so piano education can cultivate children’s sense of rhythm, guide them to establish the correct concept of sound, and deeply experience the charm of music, so that they can develop their musical aesthetic ability.

At this stage, the level of comprehensive education for young children is still low due to various complex factors such as uneven distribution of educational resources and insufficient teachers in preschool education (Guo et al., 2010). In the early childhood piano initiation education, some teachers have difficulty in linking piano learning with the cultivation of musical aesthetic ability and are obsessed with the teaching of piano playing skills, which leads to the fearfulness of young children and their inability to appreciate the beauty of music. In the kindergarten piano course, some teachers who come from professional teacher training schools are not proficient in piano themselves and lack the feeling of natural beauty in the beauty of piano music, so they cannot meet the requirements for the cultivation of musical aesthetic ability of young children (Yaseen et al., 2018). In teaching, teachers adopt an indoctrination method, requiring children to forcibly memorize various timbres, so that children can only passively follow the teacher’s rhythm to complete piano training, and cannot effectively train their perception of musical beauty (Meguro, 2020). In addition, although some teachers recognize the need to cultivate children’s musical
aesthetic ability in piano initiation education, they only educate children by playing beautiful piano music, but fail to deepen children’s experience through various aspects such as movements and expressions. For young children, it is difficult to acquire rich aesthetic feelings by listening alone due to their limited experience, and therefore, it is difficult to achieve the cultivation of musical aesthetic ability (Razidlo et al., 2018).

The traditional teaching model of piano initiation is teacher-centered, and it is mainly aimed at imparting knowledge by the teacher. Since the 1990s, the status of the teacher and the student has changed greatly in the constructivist learning environment, and thus the traditional piano initiation teaching model can no longer adapt to the modern teaching environment (Zhu et al., 2021). Multimedia Technology is a computer technology with real-time and interactivity that integrates information in the form of image, text, audio and video. The use of a multimedia piano initiation teaching system can solve the limitations and shortcomings of the traditional piano initiation teaching model and teaching methods and facilitate the application of constructivist teaching theory in piano initiation teaching. The multimedia piano initiation teaching system can be used not only for teaching various professional theory courses, but also for professional composition and piano initiation teaching creation teaching research (Peng & Xie, 2021).

At present, multimedia artificial intelligence cannot only generate accompaniment based on a given melody, but can even generate complete tracks independently, so artificial intelligence practice guidance is feasible under today’s technical conditions and is perfectly suited for more in-depth research (Shao et al., 2020). A stochastic approach using Hidden Markov Models enables the matching between score tracking and score performance to be performed. The performance deviations of three jazz pianists from mechanical regularity were objectively measured and analyzed through a MIDI-based quantization procedure to evaluate the music of these deviations. The mechanical regularity performance deviations in terms of note placement (timing), note duration (articulation) and note strength contained in 281 measures (dynamics) were used as evaluation criteria. The DTW dynamic time regularization scheme, which finds the best alignment of the track to be evaluated with the standard track, was evaluated in terms of the degree of similarity between the two-time series at the minimum distance. Although the above schemes differ in the selection of feature values, their main general idea is the same, which is to pre-process the MIDI music files first, then obtain the feature parameters, and finally compare them with the standard feature parameters to give a score based on the completeness and correctness of the performance piece (Li et al., 2021).

**Multimedia Teaching for Piano Initiation Oriented to the Cultivation of Aesthetic Ability**

Young children are full of unfamiliarity with music learning, so teachers should create a relaxing and pleasant teaching situation to attract students’ attention and stimulate their interest and enthusiasm in learning, to cultivate students’ aesthetic ability according to their psychological characteristics and growth. Introducing aesthetic teaching into music teaching can not only expand students’ knowledge, but also stimulate their interest in learning, prompting them to change from passive learners to active participants and improve their aesthetic ability (Shalmani & Branch, 2021). Applying multimedia technology to piano enlightenment teaching can break through the limitations of traditional teaching methods, create a good teaching atmosphere, effectively use classroom teaching time, stimulate students’ interest in learning, and improve students’ subjective initiative.

With the help of emotional experience, establish the aesthetic concept of music aesthetic ability in early childhood piano initiation education is the key point and the difficult point, not only can make students better memory and understanding of music knowledge, but also can promote the improvement of students’ music literacy. Therefore, teachers can use the psychology of music education to help students understand the connotation of music, understand the culture of music, and form a spiritual level through the establishment of aesthetics. In this regard, teachers can guide students to appreciate
music, and then feel the profundity of music works in the process of appreciation, so that students can better perceive the beauty of image, life, solemnity and personality expressed in music works, and then help students learn positive life education, obtain noble ideological washing, and help students improve their sound personality, enrich their spirit and soul (Zammit et al., 2020). The teacher is guided by the psychology of music education, and by grasping the focus of students’ listening, the teacher can make students immerse themselves in the situation, and then the teacher can guide students to think about the problem, which can help students better understand the connotation of musical expression, and has a very important sense of reality for the enhancement of students’ musical beauty, not only that, but also can effectively cultivate students’ appreciation ability, and can effectively promote the growth of students’ musical literacy.

Young children are still young and playful by nature, and in the process of music learning, they are easily disturbed and influenced by external factors, making it difficult for them to concentrate. Through multimedia equipment, teachers can directly and comprehensively display the teaching content to students in the form of animation, image, voice, text and other forms, so as to mobilize students’ interest and initiative in learning. In the process of interactive discussion, students can get closer to each other, enhance the understanding between students, and constantly stimulate the collision of emotions between students, which can not only improve students’ communication and expression skills, but also enable students to learn more about music interpretation, and thus develop their aesthetic skills (Tuncer & Karataş, 2022).

To develop children’s aesthetic skills in music, it is necessary to first stimulate their interest in music, so that they can listen to it more actively. Therefore, in piano initiation education, the choice of piano works should be combined with the child’s interest characteristics to fully stimulate the child’s interest and enhance the child’s musical sense (Kong, 2020). Secondly, young children are usually more active because of their age and other reasons, and it is difficult for them to listen to the piano quietly. For this situation, the teacher must create a suitable scene for the child to listen to so that the child can listen patiently and experience the emotion or feeling transmitted to the piano. In listening to music, children can discover the difference between the lunar calendar and the length and use their imagination to relate the content of the song so they can appreciate the beauty of the music. In fact, children should also focus on interest cultivation. It is obviously unscientific to practice a lot of piano skills. Therefore, teachers need to try their best to make children enjoy the fun of piano playing, with the help of beautiful notes to stimulate children to have the impulse to appreciate music, and then to meet children’s aesthetic needs (Sakaguti et al., 2019). In kindergartens, percussion activities are carried out in combination with piano performance in order to enable children to perform ensemble music collectively and separately. To this end, teachers should guide children how to use various instruments, how to follow the command gestures, and how to add to the orchestration design. In such extended activities, young children are not only better at discovering all kinds of novel things, but also master the laws of the sound emitted by such things, and cultivate their aesthetic ability to music.

METHODS

Model Structure

Automatic Music Transcription (AMT) is a fundamental problem in music information retrieval (MIR). Through AMT technology, the accuracy of instrument performance can be evaluated automatically, quickly and conveniently, so as to provide feedback for learners, so that they can quickly find problems and improve learning efficiency (Al-Radaideh et al., 2019).

In the CNN (Convolutional Neural Network)-based note level automatic transcription method, to perform feature extraction, the input sound signal is preprocessed to extract time-frequency domain features, and the spectrum is mapped to a linear space with log base using constant Q-transform to
obtain a multitone level spectral window (Sun et al., 2020; Yu & Luo, 2022). To obtain the temporal value of each note, the starting and stopping points of the notes are identified separately, both of which are performed by a separate convolutional neural network. To identify the pitch, the model uses an additional multi-pitch recognition network to identify the pitch levels near the start and stop points (Zhuo, 2021).

All the pitch levels on the note start and stop points are paired to obtain a complete information of note pitch and duration. Since the generation of note sequences goes through a note alignment process at the start and stop points, errors in the recognition of the pitch levels at the start and stop points can be a major obstacle to the alignment, and thus the scheme requires a high level of accuracy in pitch recognition. The model extracts the spectral features with CNN and then performs feature slicing to obtain the analysis using the subsequent sequence recognition layer (Suming & Bing, 2016). Since the LSTM (Long Short-Term Memory) usually used in RNN (Recurrent Neural Network) can only predict the output of the next moment based on the temporal information of the previous moment, while the output of the current moment is not only related to the previous state, but also may be related to the future state. For example, the prediction of the note start point depends not only on the previous state and the feature slice of the current input, but also may be related to the stop state that appears afterwards, and similarly, the same is true when predicting the note stop. Therefore, a bidirectional LSTM (Bi-LSTM) is used here to play the role of predicting the before-and-after relationship through the combination of the before-and-after LSTM.

This paper proposes a piano initiation teaching model based on multimedia and artificial intelligence technology for aesthetic ability development. In this paper, the hard sharing model in multi-task learning is chosen, and the shared features are triaged at some intermediate layer, with the shared module at the lower level used to compute the shared features and the private module at the

Figure 1. Method flow chart
higher level used to compute the unique features. The process of encoding the acoustic model based on CNN model proposed in this paper is shown in Figure 1.

**Note Feature Extraction Method**

The piano audio signal mainly consists of fundamental and overtone. Among them, the high audio is determined by the fundamental tone. The identification of the piano audio signal is mainly based on the detection of the fundamental tone cycle. The more common method of fundamental detection is the method of fundamental detection based on autocorrelation function, which belongs to a time domain detection algorithm and has the advantage of simple calculation process, but also has the disadvantage of errors of fundamental frequency doubling or half frequency, which can be expressed as:

\[ R_i(k) = \sum_{m=1}^{N-m} y_i(m)y_i^*(m+k) \]  

Based on equation (1) for calculation, three-level center clipping operation is required, this algorithm is a more common improvement algorithm, expressed as for equation (2). Where \( z_i(x) \) denotes the \( i \)-th frame of the audio signal after the time series) plus windowed framing, and \( C_L \) denotes the threshold level.

\[ z_i(x) < -C_L \]  

To detect the endpoints of the audio sequence \( w(t) \), mark the starting point as \( S(i) \). After performing note segmentation, the endpoints can be one-to-one corresponded with the note starting points. If \( T(i) \) denotes the fundamental period of the \( i \)-th note in the original audio, according to the short-time smoothness of the music, the \( \text{seg}(i) \) is solved for the autocorrelation function according to the long window with the window length of 4096, where the position of the first maximum autocorrelation function is the fundamental period. The ratio of the sampling frequency to the fundamental period is the fundamental frequency, and the fundamental frequency corresponds to the note.

**Improved Algorithm**

To solve the offset problem of the maximum peak, the peak ratio is evaluated and calculated by data frame panning, from which the correct peak point is obtained. The basic principle of data frame panning is to increase both the top and bottom of the selected signal period by 64, which is \( \text{seg}(i) \). Where the peak ratios \( p \) is 1.66 and 1.36, respectively, indicating that the peak ratios change after the data frames are panned. The peak ratio was found to fluctuate in the range of 1.45 through several trials, and the values were all greater than 1. If \( k \) denotes the peak ratio constant, the number of peak ratios \( p>k \), and the number of peak ratios \( p<k \), were recorded 8 times after panning, and if \( n_1<n_2 \), the first peak point was taken as the base tone cycle; if \( n_1>n_2 \), the largest peak point was taken as the base tone cycle.

Convolutional neural network is a typical algorithm of deep learning. It is a feed-forward neural network that includes convolutional computation and has a deep structure. It can learn audio data efficiently and quickly to extract the specific features of audio and has the ability of characterization learning. The computation is based on convolutional operations and pooling operations. The convolutional computation is expressed as:
In equation (3), \( x(t) \) is the input feature and \( w(t) \) is the feature mapping, which can be expressed as:

\[
s(i,j) = \sum_{m=0}^{M} \sum_{n=0}^{N} (w_{m,n}x_{i+m} + w_{n})
\]  

(4)

If \( f(x) \), \( g(x) \) in the real number field is predictable functions, the new function \( J(x) \) is the convolution of the function \( f(x) \) and \( g(x) \). The continuous convolution operation is expressed as in equation (5).

\[
J(x) = (f * g)(x) = \int -f(a)g(x-a)da
\]  

(5)

The discrete convolution operation expression is as in equation (6):

\[
J(t) = (f * g)(t) = \sum_{n=0}^{\infty} f(n)g(t-n)
\]  

(6)

In which, the convolution operation in CNN is discrete convolution, which can easily handle the problem of discrete data in real measurement.

**Aesthetic Ability Assessment**

Aesthetic ability assessment is performed using the end-to-end attention CBAM (Convolutional Block Attention Module) module as shown in Figure 2. Given an intermediate feature map, the module will generate an attention relation matrix to fuse the feature map along two independent dimensions of channel and space, a process that adaptively adjusts the features and improves the representation capability of the deep neural network. CBAM includes CAM (Channel Attention Module) and SAM (Spatial Attention Module) sub-modules, which are used for channel and spatial attention respectively.

![Figure 2. CBAM module](image-url)
This not only saves parameters and computing power, but also ensures that it can be integrated into the existing network architecture as a plug-and-play module.

The method utilizes the design idea of CBAM module and proposes a spatial feature enhancement module. The design of the spatial feature enhancement module in this paper is symmetrical to the design of the channel feature enhancement module in the previous subsection and serves to complement each other in terms of its role in feature adjustment. The detailed operation can be computed using equation.

\[ J(t) = (f^* g)(t) = \sum_{n=-\infty}^{\infty} f(n) g(t-n) \]  

(7)

Then, the two-dimensional descriptor \( \text{Avg} \) is passed through a convolutional layer with a convolutional kernel size of \( 7 \times 7 \), resulting in a feature enhancement matrix \( m \), which is normalized by a sigmoid function. The feature enhancement matrix adjusts the position of feature emphasis or suppression, which is calculated as:

\[ m = \text{Conv}_{\text{shared}}(\text{Avg}_s) = \sigma(f^{7 \times 7}(\text{Avg}_s)) \]  

(8)

Finally, the generated aesthetic and semantic feature enhancement maps can be summarized in the following equation.

\[ x_{as} = m \cdot x_s \]  

(9)

EXPERIMENTS AND RESULTS

Experiment Setup

These 8000 samples are divided into three parts, 3200 training samples, 2400 calibration samples and 2400 test samples. The training batch size was set to 100, and the generative adversarial loss, image reconstruction loss, content reconstruction loss, and style reconstruction loss were set to 1.0, 10.0, 1.0, and 1.0, respectively, with a learning rate of 0.001 and 300 iterations, and the learning rate was reduced every 40,000 iterations from the 100,000th iteration.

Experimental Results

In this experiment, the training learning rate of CNN is set to 0.001 in this paper, and the data samples are two kinds of samples, spectrum and note spectrum samples, and the CNN is trained under four ways of ELU, ReLU activation function and RMSProp (Root Mean Square prop) and Adam gradient descent method, and the test results are obtained as shown in Figure 3, Figure 4, and Figure 5. It can be seen from Figure 3, Figure 4, and Figure 5 that under the same conditions, the training results of note+ spectrum are higher than those of spectrum, and the classification accuracy is improved by 4%, 1.5%, 2% and 1.75%, respectively. It also further shows that the piano music classification method based on convolutional neural network proposed in this paper is feasible and effective.

As shown in Table 1, note sequence correctness is defined as the ratio of transcribed correct note sequences to all actual note sequences.
CONCLUSION

In early childhood piano initiation education, teachers should pay attention to the cultivation of young children’s musical aesthetic ability. Therefore, in practical teaching, works need to be reasonably selected in conjunction with children’s interests so that children’s musical perception can be strengthened. In this paper, we make a new attempt in piano initiation education for young children, not only using multimedia and other technologies for teaching assistance but hoping to solve some difficult problems in piano initiation education for young children through neural networks in the field of artificial intelligence, thereby helping piano students to discover and correct the problems of fingering and rhythm control in performance in time, and at the same time improve the aesthetic
The experimental results show that the proposed method continues to stimulate young children’s interest in piano appreciation and allows them to improve their aesthetic skills in general. In this paper, CNN is used to design the model. In the future, we can try to use the recurrent neural network and add attention mechanism.

**DATA AVAILABILITY STATEMENT**

The datasets used during the current study are available from the corresponding author on reasonable request.

**CONFLICT OF INTEREST**

Declares that he has no conflict of interest.

**FUNDING STATEMENT**

This work does not receive any funding.
REFERENCES


