



## The Transformative Impact of Artificial Intelligence on Musical Aesthetics and Aesthetic Perception in the Age of the Intersection of Art and Technology

Xiaohan Yu<sup>1</sup> and Lingling Fang<sup>2,\*</sup>

<sup>1</sup> Music School of Liaoning Normal University, Dalian, Liaoning, 116000, China

<sup>2</sup> Computer Science and Artificial Intelligence School of Liaoning Normal University, Dalian 116081, Liaoning, China

**SUMMARY:** *In the context of the era of the convergence of art and technology, traditional music education methods can not meet the current user needs, how to use artificial intelligence technology to accelerate the change of music aesthetics and aesthetic perception is an inevitable trend. This paper determines the source of this data, and uses crawler technology to obtain the data needed for the design of the artificial intelligence system, in order to complete the design of the data collection module. Randomly combining the post-deep learning neural network theory, music labeling features, to establish a music recommendation module based on LSTM-AM, in addition, the first use of the N-gram for coarse matching of music songs, in the use of DTW algorithm to further fine matching, so that the music retrieval is more accurate, and ultimately to complete the music retrieval module design. Finally, the user interaction module is designed from the aspect of music navigation and message pushing, and these four modules together constitute an artificial intelligence system for music aesthetics and aesthetic perception, and the system is discussed and analyzed in depth. The mean values of students' tests on the artificial intelligence system oriented to music aesthetics and aesthetic perception are 87.25, 89.26 and 91.15, which are higher than the mean values of the music system based on multilayer neural networks by 35.55, 39.15 and 40.95, and higher than those of the music system based on the intelligent cloud service platform by 36.2, 40.45 and 42.8, which indicates that students' tests on the artificial intelligence system oriented to music aesthetics and aesthetic perception are higher than those of the music system based on intelligent cloud service platform by 36.2, 40.45 and 42.8. Artificial Intelligence system evaluation in music aesthetics, aesthetic perception, and satisfaction performance satisfaction, to verify the practical application efficacy of the system, and to provide theoretical references for the change of music aesthetics and aesthetic perception empowered by Artificial Intelligence technology.*

**KEYWORDS:** *lstm-am; n-gram-dtw; artificial intelligence system; music aesthetics; aesthetic perception*

## 1 Introduction

Music aesthetics is a philosophical discipline that studies the art of music, which deals with the nature, function, value and aesthetics of music [1]. Music aesthetics considers the essence of music, an emotional expression [2]. This means that music is not only a combination of sounds, but also a transmission of emotions. Through composing and performing music, musicians

\*dcxyxh760623@163.com

<https://doi.org/10.65102/is2026560>

convey their emotions and thoughts to the listeners. This emotional expression can be a variety of complex emotions such as joy, sadness, anger, serenity, etc., as well as thoughts and feelings about various aspects of life, nature and society [3, 4]. Therefore, when musicians sing, they should not only pay attention to the performance of skills and timbre, but also dig deeper into the emotional connotation behind the songs, so that the emotions of the composer and lyricist can be truly conveyed to the audience. In addition, music aesthetics also emphasizes that the function of music is to meet people's aesthetic needs through emotional communication [5]. As an art form, music has a unique aesthetic value [6]. People can feel the existence of beauty in the process of appreciating music, so as to obtain spiritual pleasure and satisfaction. At the same time, music is also a kind of emotional resonance, which can bring people closer to each other and promote spiritual communication [7]. Therefore, when musicians sing, they should pay attention to the emotional interaction with the audience, infecting the audience through their songs and making them resonate, so as to achieve the purpose of emotional communication [8].

Music aesthetics is a discipline that studies music beauty and its laws, which studies both the nature, characteristics and laws of music art, as well as the generation and embodiment of music beauty [9]. Music aesthetics contains two aspects, on the one hand, it is the epistemological study of music beauty, exploring the essential attributes and characteristics of music beauty as well as the satisfaction of music beauty to the aesthetic needs of human beings [10]; on the other hand, it is the regularity study of the generation and development of music beauty, exploring the aesthetic principles that must be followed in music activities such as music creation, performance and appreciation [11]. Through these two aspects of research, music aesthetics reveals the nature of music beauty can meet the aesthetic needs of people, reflecting the aesthetic ideals of people; elucidates the aesthetic laws that must be followed in the creation, performance and appreciation of music art, in order to guide the practice of music creation, and to improve the aesthetic value and artistic effect of music art [12, 13].

Music aesthetics sets many aesthetic standards for music forms. Musical form shows wholeness and unity, and each musical element should be organically integrated to construct a harmonious and unified situation [14]. Musical form embodies the beauty of tension and release, the use of contrast, development and transition and other techniques to create a musical mood, triggering the listener's empathy [15]. The music form shows innovation and uniqueness, breaks through the old pattern, and presents a novel aesthetic perception for the listener. Music form also needs to reflect certain characteristics of the times and national cultural qualities, reflecting the aesthetic orientation of modern people and national cultural characteristics [16, 17]. Music form has a strong expressive and infectious power, can accurately convey the composer's creative intent, triggering the listener's emotional resonance and aesthetic perception.

With the development of information technology, digitalization has changed the way of music dissemination and innovation path. The use of artificial intelligence and other emerging technologies has inspired the diversified dissemination of music works and the change of creation methods [18]. In terms of music production, driven by artificial intelligence, music production and audio processing have entered a new era of intelligence, greatly changing the traditional workflow and technical methods, giving rise to a new category of "computational aesthetics" [19]. At the perceptual level, AI has a profound and comprehensive impact on music art, especially driven by personalized recommendations, which has shifted the aesthetic perception process from passive acceptance to dynamic interaction [20]. With the wide application of AI artificial intelligence in music art, in-depth thoughts and definitions are needed about its impact on music aesthetics and aesthetic perception.

In this paper, in the context of the era of the convergence of art and science and technology, the establishment of an artificial intelligence system for music aesthetics and aesthetic

perception contributes to the change and development of music aesthetics and aesthetic perception, and divides the system into four functional modules. In the data collection module, after determining the data source of this research, the crawler technology of this paper is used to obtain the system data. In the music recommendation module, a music recommendation model based on LSTM-AM is established by combining the long and short-term memory network and the two-layer attention mechanism, so as to realize the personalized recommendation of music resources, and to promote the further enhancement of users' music aesthetics and aesthetic perception. In the music retrieval module, the N-Gram index is first used to filter out most of the songs that are unlikely to be the matching results, narrowing down the selection range, and then the DTW algorithm is utilized to return the retrieved songs to the user as the results, which improves the user's experience of using the AI system oriented towards music aesthetics and aesthetic perception, and accelerates the speed of the change and development of music aesthetics and aesthetic perception. In the user interaction module, through the music navigation, message push two ways, realize the user music aesthetics and aesthetic perception intelligent control. Finally, we carry out in-depth exploration and analysis of the system from multiple dimensions in order to help the change and development of music aesthetics and aesthetic perception.

## **2 Artificial Intelligence System for Music Aesthetics and Aesthetic Perception**

By reviewing the relevant literature, it is known that the current mainstream two music systems have music platforms have a music system based on multi-layer neural networks, based on the intelligent cloud service platform of the music system of two kinds, but the performance is less than ideal. In this regard, this paper, in the context of the era of the convergence of art and science and technology, in order to explore the impact of artificial intelligence on the change of music aesthetics and aesthetic perception. Starting from the basic theory of artificial intelligence technology, an artificial intelligence system for music aesthetics and aesthetic perception is designed, which consists of a data collection module, a music recommendation module, a music retrieval module, and a user interaction module, with the aim of improving the user's music aesthetics and aesthetic perception, and then realizing the development of AI-enabled music change.

### **2.1 Data acquisition module**

Checking the research data on the artificial intelligence system for music aesthetics and aesthetic perception, it can be seen that the data in the data collection module comes from the Shrimp music website, Kmi song library, and the use of crawler technology to obtain the data required for the design of the artificial intelligence system, which consists of a system of databases, so that the system can better enhance the user's music aesthetics and aesthetic perception.

#### **2.1.1 Data sources**

In this paper, the music data comes from the music database, which is obtained from the music database crawler, including the information of all the popular singers, as well as the album information of these singers, the song information, the music data and the user data, the data information and the user's song ordering record.

### 2.1.2 Data acquisition

Getting data from the music database requires the use of crawling techniques. While crawling the data from the music database, we use depth-first search to traverse the hierarchy of artists, albums of the artists, and songs within the albums. The collected music data is in JSON format and stored in unstructured Mongo DB database. These music data include singer name, singer ID, singer heat, singer information, singer tag, album name, album ID, album heat, album information, album tag, song name, song ID, song tag.

## 2.2 Deep learning based music recommendation module

In this subsection, a recommendation algorithm with better recommendation effect is proposed based on a deep learning neural network, combining multiple music label features, fusing long and short-term memory network and two-layer attention mechanism for modeling. Compared with traditional recommendation algorithms, the algorithm refers to the user's historical data and combines it with the acoustic features of music audio, and at the same time integrates the information of songs and singers to construct the corresponding neural network model, thus realizing the personalized recommendation of music, and promoting the user's music aesthetics and aesthetic perception to be further improved.

### 2.2.1 Historical music sequence representation

A multi-LSTM network structure based on sequence temporality can use multiple LSTM units to process the user's historical data. Each LSTM unit may accept a time period of the input sequence and output a hidden state representation. These hidden state representations may be viewed as abstract representations of the user's history sequence, where each hidden state captures information about a time period in the sequence.

Specifically, the user's history can be divided into time periods, such as by each month, week, or day. Then for each time period, an LSTM unit can be used to learn the representation of that time period. In this way, at the end of each time period, the hidden state of the LSTM cell can be used as the representation  $s_h$  for that time period.

In this way, the representation  $s_h$  of each time period can capture the important information within that time period and can be updated over time. These time period representations are used to provide more detailed and accurate information when long-term preferences of the target user need to be modeled. During the training process, the historical data of the target user is used as input, and then the parameters of the LSTM network are updated by a back-propagation algorithm to predict the gap between the target and the actual observations.

First, given user  $u$ , the LSTM is used to predict all the music  $\{m_1, m_2, \dots, S_{n-1}\}$  for each historical sequence  $s_h \in \{S_1, S_2, \dots, S_{n-1}\}$  encodes the hidden layer state  $h_t$  as shown in Equation (1). Namely:

$$h_t = LSTM(x_t, h_{t-1}) \quad (1)$$

### 2.2.2 Long and Short Term Memory Networks (LSTM)

The music sequence  $S_n$  that the user interacts with is modeled using LSTM, and the user's short-term preference can be represented by the hidden state  $h_{t-1}$  of the sequence. Where the weighting operation of the current music sequence  $S_n$  is shown in equation (2). That is:

$$S_n = \frac{1}{|S_n|} \sum_{t=1}^{|S_n|} h_t \quad (2)$$

The  $S_n$  music sequence represents the user's recent preference, and after learning all the music trajectory sequences, the user's long term preference  $S_n^*$  is deduced from the current music sequence  $S_n$ , which is formally proposed to compute the user's long term preference  $S_n^*$  as shown in Eq. (3). That is:

$$S_n^* = \frac{1}{\delta(S)} \sum_k^{n-1} f(S_n, S_k) g(S_k) \quad (3)$$

The modeling process of user prediction based on long and short term preferences is shown in Figure 1.  $S_n^*$  denotes the user's long-term preference,  $h_{t-1}$  denotes the user's short-term preference,  $\delta(S) = \sum_k^{n-1} f(S_n, S_k)$  denotes the normalization factor, and  $g(\ )$  denotes the generation of  $S_k$ . To expand specifically,  $f(S_n, S_k)$  and  $g(S_k)$  are defined as shown in Eqs. (4) and (5). That is:

$$f(S_n, S_h) = \exp(S_n^T S_h) \quad (4)$$

$$g(S_h) = W_h S_h \quad (5)$$

where  $S_n$  and  $S_k$  denote the current sequence and the historical sequence, and  $W_k$  denotes the trainable weight matrix, the formula for calculating the user's long-term preference can be derived as shown in Equation (6). That is:

$$S_n^* = \frac{1}{\sum_h^{N-1} \exp(S_n^T S_h)} \sum_h^{N-1} \exp(S_n^T S_h) W_h S_h \quad (6)$$

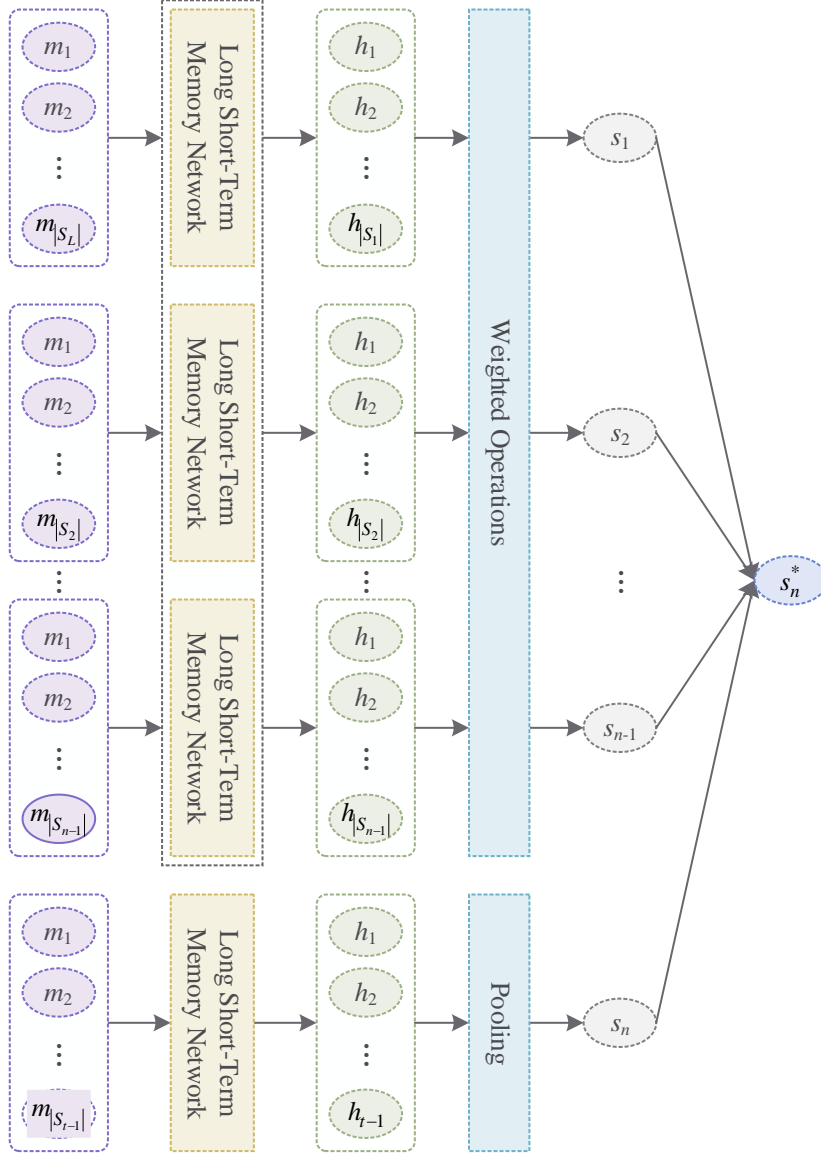


Figure 1: The LSTM modeling process

### 2.2.3 Multi-level Attention Mechanisms (AM)

Use  $m$  and  $t$  to denote the metadata features and temporal features respectively for a particular user's music history of favorable records, where the metadata features contain  $M$  modular elements, artist ID, song ID, genre, timestamp, etc., i.e.,  $m = [m_1, m_2, \dots, m_M]$ .

Firstly, all the feature vectors are mapped to get the embedded vectors of 32 dimensions, as shown in Eqs. (7) and (8). That is:

$$m'_i = W_{m_i} m_i + b_{m_i} \quad (7)$$

$$t'_i = W_t t + b_t \quad (8)$$

where  $i = 1, 2, \dots, M$ ,  $W_{m_i}$ ,  $W_t$ ,  $b_{m_i}$ ,  $b_t$  are embedding vectors. In order to distinguish the deep impact of features in different directions of music on the music preferences of different

users, a multi-layer attention mechanism is used to recognize and enhance the contribution of features in different domains, which can be divided into two levels:

The first level of attention mechanism: it is used to distinguish the importance of features in different directions to the overall music. The degree of contribution of features in different directions is determined by calculating the attention weight of each feature, so that the importance of features in different directions can be weighted according to the user's preferences and concerns.

The second level of attention mechanism: used to distinguish the impact of different users' preferences for the same feature. By calculating the attention weight of each user for the same feature, the preferences of different users are further weighted to take into account, so that the attention of different users for the same feature can be better captured and personalized recommendations can be provided. The weighted summation of the original features of music is shown in equation (9):

$$m_{att} = \sum_{i=1}^M \beta_i m'_i \quad (9)$$

where  $\beta_i$  denotes the attention score of  $m'_i$  of the embedded feature vector, and here a two-layer attention mechanism is used to compute its attention score, which is computed as shown in Equation (10). i.e:

$$\beta'_i = W_m^T \sigma(W_m m'_i + b_m) + b_m \quad (10)$$

where  $W_m \in R^{t \times d}$  and  $b_m \in R^t$  denote the parameters of the first layer,  $w_m$  denotes the parameters of the second layer, and  $t$  denotes the number of layers of the hidden layer, and the above formulas are further normalized using softmax as shown in Eq. (11). i.e:

$$\beta_i = \frac{\exp(\beta'_i)}{\sum_{i=1}^M \exp(\beta'_i)} \quad (11)$$

This finally yields the meta-feature factor  $m_{att}$  for music, and two attention networks are used in the processing of time and metadata, as they may be related to music in different ways, resulting in the music attention formula shown in (12). i.e:

$$z_{att} = \gamma_m m_{att} + \gamma_t t' \quad (12)$$

where  $\gamma_m$  and  $\gamma_t$  are the attention scores of the two features of user interaction music data features and time features respectively, the attention scores are calculated as shown in the following equation. i.e:

$$\gamma'_m = w_z^T \sigma(W_z m_{att} + b_z) + b_z \quad (13)$$

$$\gamma'_t = w_z^T \sigma(W_z t + b_z) + b_z \quad (14)$$

$$\gamma_m = \frac{\exp(\gamma'_m)}{\exp(\gamma'_m) + \exp(\gamma'_t)} \quad (15)$$

$$\gamma_t = \frac{\exp(\gamma'_t)}{\exp(\gamma'_t) + \exp(\gamma'_m)} \quad (16)$$

#### 2.2.4 LSTM-AM

To further improve the recommendation performance, an LSTM model incorporating a two-layer attentional mechanism is utilized to model the songs that have been recently praised by the user, then the definition of long and short-term memory incorporating the attentional mechanism is represented in Equation (17). i.e:

$$h'_t = LSTM(z_{att}, h'_{t-1}) \quad (17)$$

where  $h'_{t-1}$  denotes the hidden layer state. The latent factors of the user's long and short-term memory and the long and short-term memory incorporating the attentional mechanism are denoted  $h_{t-1}$  and  $h'_{t-1}$ , respectively, and the short-term preference is shown in Equation (18). To wit:

$$\tilde{h}_{t-1} = \frac{h_{t-1} * h'_{t-1}}{2} \quad (18)$$

In the music recommendation of this paper, after the LSTM model extracts the temporal features of the user's historical listening records and learns the recommendation weights through the attention mechanism, the computed and extracted temporal feature vectors are spliced horizontally with the user ID and the music ID, and this horizontal splicing operation integrates the temporal features of the user's historical listening records with the basic information of the user and the song. The splicing operation is realized by concatenating the temporal feature vectors with the user ID and music ID, or common ways such as concatenating the feature vectors, user IDs and music IDs together to form a richer feature vector can be used.

A comprehensive feature vector is obtained after horizontal splicing, which contains the temporal features of the user's historical listening records as well as the basic information of the user and the song, and the form of splicing is shown in Eq. (19). To wit:

$$\kappa = (A, \{a_1, a_2, \dots, a_n\}, B, 0 \setminus 1) \quad (19)$$

where  $\kappa$  denotes the spliced vector,  $A$  denotes the embedded multidimensional vector of user ID after data processing,  $B$  denotes the embedded multidimensional vector of item ID after data processing, and  $0 \setminus 1$  denotes whether or not the song has been interacted with the user, and if it has been interacted with, take 1, and if it has not been interacted with, take 0.

The performance and recommendation accuracy of the model can be gradually improved through continuous iterative optimization of the model parameters. After the model training is completed, the model is used to recommend new users and music, and according to the probability distribution or recommendation score calculated by the model, a personalized music recommendation list is generated for the user, which contributes to the development of the user's music aesthetics and aesthetic perception.

## 2.3 N-Gram-DTW based music retrieval module

When the number of songs in the music database is large, in order to shorten the user's waiting time, this paper adopts a two-layer matching algorithm - DTW matching retrieval algorithm based on temporal N-Gram and sentence length. The so-called two-layer matching algorithm is to divide the matching process into two times, first a coarse matching retrieval, and then a fine matching in the screening results.

(1) In the coarse matching stage, the N-Gram index is utilized to filter out most of the songs that are unlikely to be matching results and narrow down the selection. A portion of each family with the highest similarity in the coarse matching stage is selected as the range for the next matching.

(2) In the fine matching stage, due to the accuracy and effectiveness of the DTW algorithm, the DTW distance between the user's humming input and the songs in the coarse matching result is calculated one by one, and the first few songs with the smallest DTW distance are returned to the user as the result.

The music retrieval module based on N-Gram-DTW maximizes the user's experience of using the artificial intelligence system for music aesthetics and aesthetic perception, which directly affects the user's music aesthetics and aesthetic perception.

### 2.3.1 Modeling N-gram indexing

The purpose of group matching using N-Gram indexes is to filter out the songs in the database that are not likely to be the retrieval results, and to match each song in the database, so it is important to simplify the index representation and reduce the amount of computation. In this paper, we only build N-Gram sentence length index for pitch difference features, and do not consider pitch length features for the time being. The pitch change of music has a certain law, and the pitch difference between two neighboring notes is generally not more than 6 semitones. Therefore, this paper adopts the quantization criterion shown in equation (20) to represent the pitch difference sequence as a string form to improve the pitch difference tolerance of the system. Namely:

$$pitch = \begin{cases} 'Z' & (\delta\hat{x} > 6) \\ 'Y' & (4 < \delta\hat{x} \leq 6) \\ 'X' & (2 < \delta\hat{x} \leq 4) \\ 'W' & (0 < \delta\hat{x} \leq 2) \\ 'S' & (\delta\hat{x} = 0) \\ 'A' & (-2 \leq \delta\hat{x} < 0) \\ 'B' & (-4 \leq \delta\hat{x} < -2) \\ 'C' & (-6 \leq \delta\hat{x} < -4) \\ 'D' & (\delta\hat{x} < -6) \end{cases} \quad (20)$$

In the formula,  $\delta x$  represents the pitch difference between two neighboring notes, which is one of the values in the extracted pitch difference feature sequence. There is a certain limitation on the value of  $N$  in the N-Gram algorithm, it is meaningless to take too small a value, and the effect is not obvious if it is too big. The process of building the N-gram is as follows:

(1) For the pitch time series of each melody of a musical piece, calculate and obtain its pitch difference feature sequence H.

(2) Quantize  $H$  according to the criteria of the formula as containing only the characters 'Z', 'Y', 'X', 'W', 'S', 'D', 'E', 'F' and 'G' in the string  $P$ .

(3) Divide  $P$  into subsets based on sentence length.  $P = \{P_1, P_2, \dots, P_i, \dots, P_n\}$ .

(4) For each string  $E$  in  $P_i$ , divide it into strings containing only  $N$  consecutive characters, and each such string becomes an  $N$ -Gram. Append to the  $N$ -Gram its position in the string  $E$ , the position of  $E$  in the song, and put it in the database  $N$ -Gram field.

### 2.3.2 Coarse matching process

The coarse matching process utilizes matching retrieval of  $N$ -Gram indexes in the music database to obtain a collection of fine matching candidates. The specific process is as follows:

(1) For the humming speech feature sequence  $W$ , the length is  $n$ , denoted as  $|W|=n$ . According to the quantization criterion of the formula,  $W$  is converted into a string sequence  $Q$ .

(2) Generate the set  $G$  of all  $N$ -grams of  $Q$ , where  $(3 \leq N \leq 4)$ .

(3) Denote any  $x$ -gram in  $G$  as  $g$ , where  $(3 \leq x \leq 4)$ . That is, the string in  $P_i$  is  $E$ . The information of  $g$ , the position of  $g$  in  $E$ , the position of  $E$  in the song, and the sequential number of the song, which together form the set  $J$ , i.e., adding timing information to  $N$ -gram.

(4) For any two  $x$ -gram  $g_1$  and  $g_2$  in  $G$ ,  $(3 \leq N \leq 4)$ , if there also exist  $g_1$  and  $g_2$  in the set  $J$  and they have similar positional relations in  $E_x$ , the value of similarity between the two sequences is added to one.

(5) Sort  $sim(Q, E_x)$  in descending order. The serial numbers of the first  $M$  songs, and the position in the song where  $E_x$  is located, are deposited into the set  $K$ .  $K$  is a candidate set for use in the fine matching process.

### 2.3.3 Fine matching process

After the above coarse matching process, a collection of candidate songs  $K$  is obtained. Then the fine matching algorithm is run for the melody corresponding to the  $x$ th phrase of each song in the set to calculate the similarity.

Let the pitch feature sequence of the user's humming input be  $HP = \{hp_1, hp_2, \dots, hp_m\}$ , and the pitch length feature sequence be  $HR = \{hr_1, hr_2, \dots, hr_m\}$ , the length of these two sequences is  $m$ ; in candidate set  $K$ , the melodic pitch feature sequence corresponding to the  $x$ th phrase of the song is  $DP = \{dp_1, dp_2, \dots, dp_n\}$ , and the tone length feature sequence is  $DR = \{dr_1, dr_2, \dots, dr_n\}$ , and its length is  $n$ . In order to compare the similarity between them, the DTW distances  $D(HR, DR)$  and  $D(HP, DP)$  can be computed, with smaller distances implying higher similarity. As an example, the algorithm for calculating the DTW distance for  $D(HP, DP)$  is as follows:

Initial condition:  $D(1,1) = 0$ ;

Recursion:  $D(i, j) = d(i, j) + \min(D(i-2, j-1), D(i-1, j-1), D(i-1, j-2))$ ;

Termination:  $D_{\min}(HP, DP) = D(m, n)$ .

In the above equation,  $i$  denotes the pitch value of the  $i$ th note of the humming feature

sequence, and  $j$  denotes the pitch value of the  $j$ th note of the  $x$ th phrase of the song in the candidate set.  $d(i, j) = |i - j|$  denotes the Euclidean distance between two pitch values.  $D_{\min}(HP, DP)$  is the distance between two pitch time series, where a smaller distance indicates a higher similarity between the two feature sequences.

Finally, the values of pitch similarity and rhythmic similarity are combined and compared to get the similarity between the humming input and the phrase. The formula for calculating the similarity by combining the pitch and rhythm DTW values is as follows:

$$Sim = 100 - (\beta \cdot D_{\min}(HP, DP) + (1 - \beta) \cdot D_{\min}(HR, DR)) \quad (21)$$

where  $\beta$  is a weight, set by yourself. The  $\beta$  indicates how important the pitch is relative to the length of the note in reflecting the melody of the song.  $\beta$  is a ratio, so  $\beta \in [0, 1]$ . Since most people are better at grasping pitch than pitch length, pitch plays a greater role in the humming information, and a value of  $\beta$  greater than 0.5 is preferred.

In the process of fine matching, the humming feature sequence is compared with the first  $x$  sentence of each song in the candidate set  $K$ , and there may be more than one sentence that meets the conditions of coarse matching in some songs, and at this time, the similarity of the melody of the sentence with the largest similarity is returned as the similarity of the song. Finally, the songs are sorted in descending order of similarity, and a number of songs are returned as retrieval results.

## 2.4 User interaction module

User interaction module is a module designed to facilitate the user to use the system, the design principle is to let the user input as little information as possible, as much as possible to present information to the user, the interface is simple and clear, the operation is simple and easy to use. In this paper, the interaction between the system and the user includes two kinds of music navigation and message pushing to realize the intelligent control of the user's music aesthetics and aesthetic perception.

### 2.4.1 Music navigation

The music navigation function consists of three main parts: song navigation, artist navigation, and search. The main page of song navigation is a list of songs sorted by song heat, and the optional navigation items are song style, mood, theme, and scene; the main page of singer navigation is a list of singers sorted by singer heat, and the optional navigation items are gender, region, and style; the retrieval brings convenience for users to quickly find, and the page will carry out different types of data feedback according to the retrieval results, so as to provide the users with the development of music aesthetics and aesthetic perception. The page will provide different types of data feedback according to the search results, so as to escort the development of users' music aesthetics and aesthetic perception.

### 2.4.2 Message Push

Message push is a personalized service function added to enrich the system function, in order to achieve this function, the system needs to regularly grab the news dynamics of each singer from the entertainment news website and store it in the MySQL database, when the user uses the system, the system takes out the news dynamics of the singer that the user likes from the nwes table in the database and pushes it to the user, so as to make the user reach a new level of

music aesthetics and aesthetic perception. This will make the user's music aesthetics and aesthetic perception reach a new height.

### 3 Systematic Exploratory Analysis

#### 3.1 System module testing

In order to ensure the validity of the test results of the artificial intelligence system for music aesthetics and aesthetic perception, the system test environment should be built as close as possible to the actual operating environment of the system. The clients of the system are PC and cell phone, and the test environment of the PC side of the system should contain browsers with three kernels: IE kernel, Google kernel and Firefox kernel, while the cell phone side should contain browsers of the two mainstream cell phone platforms, Android and IOS. In the actual operating environment, the use of LINUX operating system can obtain higher security and stability, so all servers in the test environment use LINUX operating system. The purpose of the performance test is to verify whether the AI system for music aesthetics and aesthetic perception meets the performance design requirements, can satisfy the user experience, and ensure the smooth operation of the AI system for music aesthetics and aesthetic perception.

##### 3.1.1 Data Acquisition Module Testing

After the system is developed, it needs to be stress-tested to verify whether the program meets the design requirements. Web performance testing has the following performance indicators: transaction response time, the number of concurrent users, throughput, the number of transactions per second (TPS), the click-through rate and resource utilization. In this paper, we use LoadRunner software to focus on the average transaction response time and concurrent access to the system during the server-side stress test. Test process concurrent access from 20 to start, each increase of 20 concurrent access, each stage of the pressure lasts a few minutes and statistics of the average response time of the system at this stage, the data collection module test results are shown in Table 1. Through the data performance in the table, it can be seen that the average response time of the system increases slowly with the increase of concurrent accesses, and meets the design requirement of response time less than 5 seconds when the concurrent accesses are less than 1,000, which indicates that the data acquisition module in the artificial intelligence system for music aesthetics and aesthetic perception can operate normally and thus promote the development of users' music aesthetics and aesthetic perception.

*Table 1: Test results of the data acquisition module*

User	Average response time/s	User	Average response time/s	User	Average response time/s	User	Average response time/s	User	Average response time/s
20	0.038	220	0.918	420	2.122	620	3.203	820	4.392
40	0.159	240	0.952	440	2.153	640	3.234	840	4.438
60	0.173	260	1.041	460	2.322	660	3.366	860	4.481
80	0.231	280	1.048	480	2.41	680	3.574	880	4.55
100	0.275	300	1.22	500	2.476	700	3.63	900	4.634
120	0.401	320	1.416	520	2.592	720	3.727	920	4.707
140	0.487	340	1.536	540	2.603	740	4.24	940	4.784
160	0.537	360	1.756	560	2.721	760	4.276	960	4.833
180	0.575	380	1.982	580	2.846	780	4.361	980	4.968
200	0.602	400	2.025	600	3.199	800	4.368	1000	4.985

### 3.1.2 Music recommendation module testing

#### (1) Algorithm testing

The datasets 1K-users and MIGU are selected as the data sources for testing the LSTM-AM combination algorithm, and the proposed LSTM-AM combination algorithm is compared with the following generalized deep-learning based recommendation methods. Namely:

**LSTM:** This is a variant of the RNN model that performs efficiently in handling sequential data.

**DeepFM:** It combines a factorization machine and a neural network.

In addition, the following music-oriented recommendation methods are considered. Namely:

**c-IFRM:** a hybrid music recommendation method that combines clustering and latent factor modeling for users' music play records.

**DCUE:** short for Deep Content-User Embedding Model, a hybrid approach that utilizes user-item interaction and music audio content.

**MEM:** Modeling user's global and contextual music preferences through music embedding based on their listening history.

To evaluate the accuracy of the LSTM-AM and comparison algorithms, the 1K-users and MIGU datasets are trained and tested, respectively, and the recall and NDCG metrics are used to measure the recommendation quality. In order to eliminate the instability of individual experimental results, this paper repeats 20 experiments, and the average of the experimental results is used as the final evaluation result. The recommendation performance analysis of different methods is shown in Fig. 2, where (a)~(b) are the datasets 1K-users and MIGU, respectively. On the two experimental datasets, the LSTM-AM proposed in this paper consistently and significantly outperforms all the comparison models. On the 1K-users dataset, Recall@20 and NDCG@20 reached 0.868 and 0.866, respectively. On the MIGU dataset, Recall@10 and NDCG@20 reached 0.881 and 0.902 respectively. The quantitative evaluation results clearly demonstrated the superiority of the LSTM-AM algorithm in the music recommendation module.

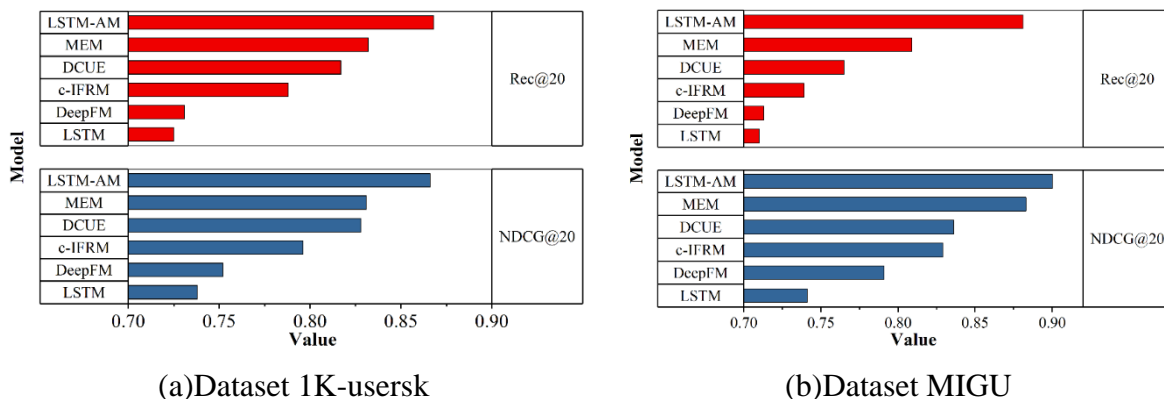


Figure 2: Analysis of the recommendation performance of different methods

#### (2) Response time test

After verifying the superiority of the LSTM-AM algorithm on the music recommendation module, the overall response time of the music recommendation module is then tested, and the test results of the music recommendation module are shown in Table 2. After testing, it can be seen that the average response time of the music recommendation module grows with the number of concurrencies, and the overall monotonically increasing trend, when the number of system concurrencies reaches 1,000, the average response time of the music recommendation module rises to 4.949s, and does not exceed 5s, which indicates that the music recommendation

module based on the combined LSTM-AM algorithm has excellent response efficiency, and is able to give the user with a quality experience, so that the user's music aesthetics and aesthetic perception is better developed.

Table 2: Test results of the music recommendation module

User	Average response time/s	User	Average response time/s	User	Average response time/s	User	Average response time/s	User	Average response time/s
20	0.179	220	1.408	420	2.13	620	3.253	820	4.166
40	0.239	240	1.504	440	2.195	640	3.48	840	4.224
60	0.257	260	1.524	460	2.336	660	3.729	860	4.465
80	0.288	280	1.531	480	2.435	680	3.886	880	4.473
100	0.721	300	1.629	500	2.458	700	3.886	900	4.535
120	0.933	320	1.653	520	2.54	720	3.978	920	4.564
140	1.075	340	1.767	540	2.609	740	4.009	940	4.57
160	1.122	360	1.884	560	2.811	760	4.118	960	4.665
180	1.162	380	1.955	580	2.903	780	4.144	980	4.675
200	1.32	400	2.02	600	2.977	800	4.162	1000	4.949

### 3.1.3 Music Search Module Test

#### (1) Algorithm testing

At present, the music retrieval technology is still in the stage of exploration and research, and there is no unified music database and detection standard to test the performance of the algorithm, so in this paper, 10 pieces of popular music (A, B, C, D, E, F, G, H, I, J) are randomly collected from the Internet as the test object of the music retrieval algorithm. In order to be able to better reflect the superiority of the N-Gram-DTW algorithm, the algorithm is analyzed with the N-Gram algorithm and the DTW algorithm for music retrieval accuracy, and the analysis of the algorithm test results is shown in Figure 3. Based on the data performance in the figure, it can be seen that the N-Gram-DTW algorithm is superior to the N-Gram algorithm and the DTW algorithm in terms of music retrieval accuracy, and the distribution interval of music retrieval accuracy is 0.853~0.942, which fully verifies the priority of the N-Gram-DTW algorithm in the music retrieval module, and thus ensures that the module of the artificial intelligence system oriented towards music aesthetics and aesthetic perception is Feasibility of design.

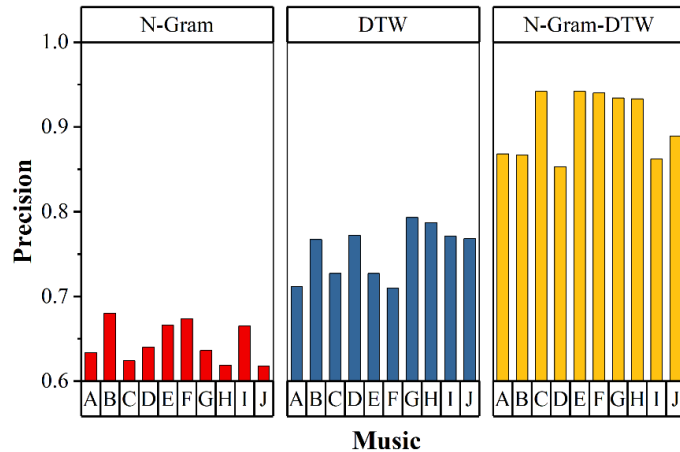


Figure 3: Analysis of algorithm test results

## (2) Response time test

After detecting the priority of the N-Gram-DTW algorithm in the music retrieval module, the music retrieval module as a whole is then tested, and the average response time test results are shown in Table 3. The size of the data in the table shows that the average response time of the music retrieval module as a whole shows a monotonically increasing trend with the number of concurrencies, and as the number of concurrencies rises from 20 to 1000, the average response time of the music retrieval module as a whole also rises from 0.037s to 4.971s, and no matter how the average response time of the music retrieval module as a whole grows, it is always maintained within 5s, which fulfills the standard requirements for the response time test of the system module. The average response time of the music retrieval module always stays within 5s, which meets the standard requirements of the system module, indicating that the music retrieval module in the system is able to bring users a high-quality experience, which maximizes the users' music aesthetics and aesthetic perceptions.

*Table 3: Average response time test results*

User	Average response time/s	User	Average response time/s	User	Average response time/s	User	Average response time/s	User	Average response time/s
20	0.037	220	0.897	420	1.76	620	2.993	820	4.11
40	0.075	240	0.927	440	1.787	640	3.019	840	4.111
60	0.094	260	0.974	460	1.82	660	3.19	860	4.572
80	0.122	280	1.057	480	1.92	680	3.263	880	4.606
100	0.188	300	1.119	500	1.987	700	3.306	900	4.615
120	0.301	320	1.127	520	2.301	720	3.326	920	4.703
140	0.427	340	1.461	540	2.599	740	3.509	940	4.852
160	0.436	360	1.486	560	2.686	760	3.604	960	4.874
180	0.585	380	1.587	580	2.884	780	3.857	980	4.937
200	0.661	400	1.643	600	2.9	800	3.897	1000	4.971

### 3.1.4 User interaction module testing

After testing the data acquisition module, music recommendation module, and music retrieval module in the artificial intelligence system for music aesthetics and aesthetic perception, the user interaction module is tested next, and the results of the user interaction module are shown in Table 4. Based on the data performance in the table, it can be seen that the average response time of the artificial intelligence system for music aesthetics and aesthetic perception shows a monotonically increasing trend with the number of concurrent accesses, and when the number of concurrent accesses is increased to 1,000, the average response time of the system is 4.952s, which can be derived from the fact that the user interaction module meets the design requirement of a response time of less than 5 seconds under the case of the concurrent accesses lower than 1,000, which confirms that the user interaction module for music aesthetics and aesthetic perception has been tested. The stability of the user interaction module for music aesthetics and aesthetic perception, which in turn realizes the intelligent control of users' music aesthetics and aesthetic perception, and also accelerates the change of music aesthetics and aesthetic perception.

Table 4: Test results of the user interaction module

User	Average response time/s	User	Average response time/s	User	Average response time/s	User	Average response time/s	User	Average response time/s
20	0.198	220	1.372	420	1.987	620	3.067	820	4.183
40	0.265	240	1.456	440	2.235	640	3.413	840	4.263
60	0.351	260	1.465	460	2.365	660	3.535	860	4.279
80	0.433	280	1.5	480	2.524	680	3.551	880	4.304
100	0.53	300	1.507	500	2.527	700	3.696	900	4.32
120	0.541	320	1.579	520	2.638	720	3.884	920	4.333
140	0.545	340	1.643	540	2.779	740	3.938	940	4.501
160	0.562	360	1.849	560	2.862	760	4.106	960	4.8
180	0.797	380	1.9	580	2.958	780	4.123	980	4.919
200	1.001	400	1.929	600	3.052	800	4.126	1000	4.952

## 3.2 System application performance testing

### 3.2.1 Test preparation

Referring to the existing research content, we can see that the two mainstream music platforms have a multi-layer neural network-based music system and an intelligent cloud service platform-based music system as a comparison system, which are named A and B, aiming at reflecting the priority of the music aesthetics and aesthetic perception oriented AI system, and naming the proposed music aesthetics and aesthetic perception oriented AI system as C. Considering that the usage paths of the target users are not the same, such as teachers' usage paths for teaching arrangement, teaching design, etc., while students' usage paths are more for listening to music. Considering the inconsistency of the usage paths of the target users, such as teachers' usage paths of teaching arrangement and teaching design, and students' usage paths of listening to music, 20 teachers and 20 students are selected as the test users of the system's application efficacy. From the three aspects of music aesthetics, aesthetic perception, and satisfaction, the evaluation scores were assigned, with poor, poor, average, good, and good scores ranging from 0 to 20, 20 to 40, 40 to 60, 60 to 80, and 80 to 100, respectively.

### 3.2.2 Test results

#### (1) Teachers

Twenty teachers were invited to test and rate the application efficacy of the three different systems from the three aspects of music aesthetics, aesthetic perception, and satisfaction, and the results of the teachers' test of the application efficacy of the systems are shown in Table 5. From the distribution of data in the table, it can be seen that in the aspect of music aesthetics, the average score of teachers' evaluation of AI systems oriented to music aesthetics and aesthetic perception is 91.4, while the average score of teachers' evaluation of music systems based on the intelligent cloud service platform and music systems based on multilayer neural networks is 50.85 and 49.6, that is to say, it is verified that AI systems oriented to music aesthetics and aesthetic perception are prioritized, and the same reason exists in the aspects of aesthetic perception, satisfaction in the presence of synesthesia, demonstrating that the system can effectively meet the needs of teachers in the use of instructional design, teaching programs, and thus help music aesthetics in the development of aesthetic perception.

*Table 5: The teacher's test results on the application effectiveness of the system*

Teacher	Music aesthetics			Aesthetic perception			Satisfaction		
	A	B	C	A	B	C	A	B	C
1	50	48	94	42	46	83	42	48	81
2	42	55	95	49	45	89	42	51	84
3	52	47	100	43	50	81	57	52	84
4	50	58	89	54	60	84	55	43	83
5	45	47	85	48	45	93	51	54	95
6	59	47	97	55	53	90	42	41	95
7	56	46	84	58	40	95	54	55	87
8	50	60	90	54	53	86	44	45	94
9	60	53	82	44	58	96	57	51	94
10	44	53	95	51	50	86	54	51	87
11	45	60	85	45	57	85	43	40	84
12	42	47	97	47	44	88	54	53	80
13	52	46	97	41	52	87	51	57	90
14	40	46	92	48	56	81	50	55	82
15	59	59	99	45	50	99	48	57	90
16	43	53	87	44	48	86	55	42	84
17	41	55	91	50	52	99	54	50	94
18	59	46	91	45	44	90	56	55	98
19	47	48	95	54	52	96	59	42	91
20	56	43	83	54	40	86	55	55	87
Mean	49.6	50.85	91.4	48.55	49.75	89	51.15	49.85	88.2

## (2) Students

In order to make the above findings more convincing, 20 students were also invited to test the application efficacy of different systems from the three aspects of music aesthetics, aesthetic perception, and satisfaction, and the results of the students' evaluation of the application efficacy of different systems are shown in Table 6. It can be seen from the data performance in the table that the artificial intelligence system oriented to music aesthetics and aesthetic perception has mean values of 87.25, 89.6, and 91.15 in the three aspects of music aesthetics, aesthetic perception, and satisfaction, and the music system based on multilayer neural network has mean values of 51.7, 50.45, and 50.2 in the three aspects of music aesthetics, aesthetic perception, and satisfaction, and the music system based on intelligent cloud service platform has mean values of 51.7, 50.45, and 50.2 in the three aspects of music aesthetics, aesthetic perception, and satisfaction. The mean values of the music system based on intelligent cloud service platform in the three aspects of music aesthetics, aesthetic perception, and satisfaction are 51.05, 49.15, and 48.35, which can be concluded that the students are more optimistic about the effectiveness of the application of the artificial intelligence system oriented towards music aesthetics and aesthetic perception, and it can satisfy the students' needs for listening to music, music aesthetics, and aesthetic perception, and it can all-around validate the effectiveness of the application of the artificial intelligence system oriented towards music aesthetics and aesthetic perception. This is a comprehensive verification of the application effect of the artificial intelligence system for music aesthetics and aesthetic perception.

Table 6: The evaluation results of students' application performance of different systems

Students	Music aesthetics			Aesthetic perception			Satisfaction		
	A	B	C	A	B	C	A	B	C
1	59	46	89	46	45	92	58	43	98
2	52	44	82	58	52	93	50	50	88
3	49	57	88	46	55	81	48	44	100
4	52	50	90	44	41	81	53	45	96
5	51	48	91	45	40	88	47	48	81
6	60	49	82	57	56	98	57	46	96
7	52	41	94	57	59	84	44	50	99
8	52	57	84	58	42	90	52	43	90
9	43	57	83	50	52	90	44	40	96
10	59	51	91	57	43	100	60	51	93
11	57	59	80	42	51	83	50	51	83
12	47	46	82	49	54	98	59	53	99
13	51	57	87	42	42	94	45	45	81
14	58	40	87	53	45	86	46	58	94
15	42	55	83	50	49	85	56	58	96
16	60	51	89	43	41	94	59	54	89
17	41	47	81	51	42	94	40	40	81
18	52	47	100	57	60	85	43	58	98
19	47	59	91	48	55	81	46	46	84
20	50	60	91	56	59	95	47	44	81
Mean	51.7	51.05	87.25	50.45	49.15	89.6	50.2	48.35	91.15

### (3) Mean value

In order to better visualize the results of the teachers' and students' tests of the effectiveness of different system applications, the above teachers' and students' test means of the system were statistically summarized, and the summary of the teachers' test means of the effectiveness of different system applications is shown in Table 7, and the summary of the students' test means of the effectiveness of different system applications is shown in Table 8. The mean values of teachers' evaluation of the artificial intelligence system oriented to music aesthetics and aesthetic perception in the three aspects of music aesthetics, aesthetic perception, and satisfaction are 91.4, 89, and 88.2, which are higher than those of the music system based on multilayer neural network by 41.8, 40.45, and 37.05, and higher than those of the music system based on the intelligent cloud service platform by 40.55, 39.25, and 38.35, which indicates that teachers hold a better and good attitude towards the performance of the application efficacy of the artificial intelligence system oriented to music aesthetics and aesthetic perception. And the mean values of students' evaluation of the artificial intelligence system oriented to music aesthetics and aesthetic perception are 87.25, 89.26, and 91.15, which are 35.55, 39.15, and 40.95 higher than the mean values of the music system based on multilayer neural network, and 36.2, 40.45, and 42.8 higher than the mean values of the music system based on intelligent cloud service platform, which indicates that students fully affirm the performance of the artificial intelligence system oriented to music aesthetics and aesthetic perception, which indicates that students fully affirm the performance of the artificial intelligence system oriented to music aesthetics and aesthetic perception. Aesthetic Perception of Artificial Intelligence System Evaluation in the application of music aesthetics, aesthetic perception, and satisfaction performance. In this paper, by constructing an artificial intelligence system for music aesthetics

and aesthetic perception, we promote the change and development of music aesthetics and aesthetic perception in the context of the era of the convergence of art and technology.

*Table 7: The average value of teachers' application efficiency of the system*

Teacher	A	B	C
Music aesthetics	49.6	50.85	91.4
Aesthetic perception	48.55	49.75	89
Satisfaction	51.15	49.85	88.2

*Table 8: The average value of students' application efficiency of the system*

Students	A	B	C
Music aesthetics	51.7	51.05	87.25
Aesthetic perception	50.45	49.15	89.6
Satisfaction	50.2	48.35	91.15

## 4 Conclusion

In the context of the era of convergence of art and technology, how to use artificial intelligence technology to accelerate the change and development of music aesthetics and aesthetic perception has become a hot topic in the field of music research. In this paper, we refer to the existing research results and construct an artificial intelligence system for music aesthetics and aesthetic perception to help the change and development of music aesthetics and aesthetic perception in the context of the era of convergence of art and technology.

(1) In the module test of the artificial intelligence system for music aesthetics and aesthetic perception, the average response time shows a monotonically increasing trend with the number of concurrency, and when the number of concurrency grows to 1,000, the average response time still remains within 5s, which meets the standard requirement of the average response time of the system, which indicates that the various modules of the artificial intelligence system for music aesthetics and aesthetic perception have a stable operation state, so that it can promote the development of Change and development of music aesthetics and aesthetic perception

(2) Teachers' evaluation of the artificial intelligence system for music aesthetics and aesthetic perception in the three aspects of music aesthetics, aesthetic perception, and satisfaction averaged 91.4, 89, and 88.2, and was more excellent than the music system based on multi-layer neural network and the music system based on intelligent cloud service platform, which fully demonstrates the effectiveness of the application of the artificial intelligence system for music aesthetics and aesthetic perception to make the art and technology The change of music aesthetics and aesthetic perception in the context of the era of convergence is further developed.

## References

- [1] Downes, S. (Ed.). (2014). *Aesthetics of Music: Musicological Perspectives*. Routledge.
- [2] Wu, H. (2025). The Aesthetics of Musical Expression: A Philosophical Exploration of Emotion and Meaning in Music. *Mediterranean Archaeology and Archaeometry*, 25(3).
- [3] Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in psychology*, 4, 487.

- [4] Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 921.
- [5] Jakubowski, K., Polak, R., Rocamora, M., Jure, L., & Jacoby, N. (2022). Aesthetics of musical timing: Culture and expertise affect preferences for isochrony but not synchrony. *Cognition*, 227, 105205.
- [6] Fingerhut, J., Gomez-Lavin, J., Winklmayr, C., & Prinz, J. J. (2021). The aesthetic self. The importance of aesthetic taste in music and art for our perceived identity. *Frontiers in psychology*, 11, 577703.
- [7] Tong, F. (2024). Human motion and musical expression: Exploring the link between physical movement and emotional resonance in performance. *Molecular & Cellular Biomechanics*, 21(2).
- [8] Dash, A., & Agres, K. (2024). Ai-based affective music generation systems: A review of methods and challenges. *ACM Computing Surveys*, 56(11), 1-34.
- [9] Martin, R., & Nielsen, N. (2024). Enacting musical aesthetics: The embodied experience of live music. *Music & Science*, 7, 20592043231225732.
- [10] Mehl, K., Gugliano, M., & Belfi, A. M. (2023). The role of imagery and emotion in the aesthetic appeal of music, poetry, and paintings. *Psychology of Aesthetics, Creativity, and the Arts*.
- [11] Clemente, A., Pearce, M. T., & Nadal, M. (2022). Musical aesthetic sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, 16(1), 58.
- [12] Braun Janzen, T., de Oliveira, B., Vantorim Ferreira, G., Sato, J. R., Feitosa-Santana, C., & Vanzella, P. (2023). The effect of background music on the aesthetic experience of a visual artwork in a naturalistic environment. *Psychology of Music*, 51(1), 16-32.
- [13] Li, Y., & Sun, R. (2023). Innovations of music and aesthetic education courses using intelligent technologies. *Education and Information Technologies*, 28(10), 13665-13688.
- [14] Juslin, P. N., Ingmar, E., & Danielsson, J. (2023). Aesthetic judgments of music: Reliability, consistency, criteria, self-insight, and expertise. *Psychology of Aesthetics, Creativity, and the Arts*, 17(2), 193.
- [15] Czepiel, A., Fink, L. K., Seibert, C., Scharinger, M., & Kotz, S. A. (2023). Aesthetic and physiological effects of naturalistic multimodal music listening. *Cognition*, 239, 105537.
- [16] Wang, H., & Sourin, A. (2023). Visual signatures of music mood. In *SIGGRAPH Asia 2023 Posters* (pp. 1-2).
- [17] Kurbanbaevich, Z. K. (2025). The musical and aesthetic culture of the individual: essence and components. *International Journal of Pedagogics*, 5(02), 27-28.
- [18] Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM transactions on multimedia computing, communications, and applications*

(TOMM), 18(2), 1-22.

- [19] Wei, L., Yu, Y., Qin, Y., & Zhang, S. (2025). From Tools to Creators: A Review on the Development and Application of Artificial Intelligence Music Generation. *Information*, 16(8), 656.
- [20] Sato, M., & McKinney, J. (2022). The enactive and interactive dimensions of ai: Ingenuity and imagination through the lens of art and music. *Artificial life*, 28(3), 310-321.