Hindawi
Advances in Multimedia
Volume 2022, Article ID 6200077, 13 pages
https://doi.org/10.1155/2022/6200077

Research Article

Multimedia Pop Music Teaching Model Integrating Semifinished Teaching Strategies

Kangtan Dong [1,2]

[1] *Music School, Xinyang College, Xinyang 464000, China*
[2] *Xinyang Normal University, Xinyang 464000, China*

Correspondence should be addressed to Kangtan Dong; dkt@xynu.edu.cn

Received 23 November 2021; Revised 10 December 2021; Accepted 23 December 2021; Published 22 January 2022

Academic Editor: Qiangyi Li

In order to improve the effect of popular music teaching, this paper combines the intelligent music frame feature recognition technology and the semifinished product teaching strategy to construct a multimedia popular music teaching system. Moreover, this paper reexpresses energy detection and zero-crossing detection mathematically from the perspective of aggregation and proposes a new two-level detection method for music frames. In addition, this paper uses the fusion and expansion of the zero-crossing rate detection result to judge whether a piece of speech is a valid musical speech. Finally, this paper uses a four-tier architecture to design the system framework, constructs the system functional modules, builds a multimedia pop music teaching system that incorporates semifinished teaching strategies, and then validates the performance of the system. From the experimental research, it can be seen that the multimedia pop music teaching system integrated with the semifinished product teaching strategy proposed in this paper can play an important role in music teaching and effectively improve the effect of music teaching.

1. Introduction

As a kind of music culture with great characteristics of the times, popular music is more and more widely used in music teaching, and it has won the attention of all walks of life. Diversification is the right direction for the development of art. As a kind of art, the types of music should also be diversified. Moreover, music teaching should focus on not only traditional music, but also modern music [1]. Popular music meets the aesthetic and psychological needs of the public, and the appreciation of popular music requires the audience to have sufficient imagination and aesthetic ability. There are countless popular music works today. Music teachers need to pay attention to popular music that is closer to students' lives and has a profound impact on students and introduce it into music teaching reasonably, so that students can learn popular music more happily, understand popular music more deeply, feel the beauty of popular music more fully, and then create an efficient music classroom [2].

Once, people classified "pop music" as popular, simple, and clear popular music and also generally referred to "pop songs" with short structure. Once upon a time, it has become synonymous with youth "rebellion." Moreover, many music scholars have their own opinions and opinions on whether "pop music" can appear in various professional colleges and establish separate disciplines. Since the reform and opening up, with the continuous development of the national economy, national culture, and entertainment living standards, our understanding of popular music has been updated and deepened [3].

We spend more time understanding and studying the development history of popular music and overturning outdated and narrow cognitions in a gradual understanding. Since the birth of Jazz in the 19th century, this nascent art form has incorporated more forms and possibilities and has integrated cultures from different regions in a century of development. The cultures of many countries and different nationalities are constantly developing and progressing, and people are willing to learn and explore for them. In fact, any form of musical expression and musical style contains abundant regional culture and national characteristics. More music scholars will integrate their personal emotions and absorb the best of the culture of various countries, and show them in the form of "auditory". Several famous music schools in the world have established professional disciplines for popular music, such as the Manhattan School of Music and Berkeley School of Music in the United States and the Royal Academy of Music in the United Kingdom. These schools not only have achieved brilliance in various music systems, but are also at the world's advanced level in the form and diversity of popular music, with mature forms and sound systems. In recent years, major music schools in China have also made active efforts in the subject of popular music and have achieved certain developments. In addition to the establishment of special disciplines, a large number of pop music talents have also been cultivated, and they have emerged and are active on the music stage.

Based on this, this paper integrates the teaching strategies of semifinished products to construct a multimedia pop music teaching system and evaluates the teaching effect of the system, so as to provide a theoretical reference for the follow-up pop music teaching.

2. Related Work

Literature [4] discussed three aspects: the historical evolution of music communication media, the emergence of new audiovisual media, and the dominant trend of online music communication. Literature [5] sorted out and analyzed the content of modern educational technology. In the aspects of informatization teaching design and informatization teaching practice, literature [6] has made a more detailed elaboration. Literature [7] analyzed the importance of teachers' information and communication technology and collaboration ability requirements in teachers' professional competence standards by comparing with the old standards. Literature [8] elaborated on the problems existing in music appreciation teaching and singing teaching and pointed out that music teaching should not deviate from the music itself and should use music as the main line to guide students to experience the image and content of music under the stimulation of sound. Literature [9] pointed out that, in music classroom teaching, it is necessary to reflect the basic concept of the curriculum and to not only highlight the characteristics of the subject, but also pay attention to the integration of the subject. Literature [10] explains the current misunderstandings in primary school music teaching from five aspects. It includes the misunderstanding that the implementation of the new curriculum standards is not in place, the misunderstandings of emphasizing the form of the classroom and not the transfer of knowledge, the misunderstandings of focusing on individual training and ignoring overall improvement and development, the misunderstandings of emphasizing textbook teaching and ignoring the development of local music curriculum, and the misunderstanding of emphasizing the improvement of teaching conditions and ignoring the full use of teaching conditions.

Literature [11] pointed out that music teaching should arouse students' desire for knowledge and true feelings and promote students' aesthetic ability. Literature [12] mainly discussed the hot and difficult issues in current music education. Literature [13] comprehensively analyzed the principles, learning modes, and implementation methods in the integration process of information technology and music courses on the basis of existing theoretical analysis. Literature [14] analyzed the pros and cons of multimedia in music education in primary and secondary schools. According to the current application of multimedia technology to music teaching, literature [15] has made some thoughts and studies on the role and prospects of multimedia in the actual process of music education through the breadth and depth of multimedia application and the perspective of teaching achievements. Literature [16] proposed that the correct use of multimedia teaching methods for music teaching is of great significance to the improvement of music teaching theories and standards. Literature [17] studied people's activities in music education from the perspective of psychology, including the psychological activities that students find when experiencing music, expressing music, creating music, and learning music knowledge and skills and music culture and psychological activities during music appreciation, which are used to guide educational activities. This has provided a great help for the research in the music classroom of elementary and middle schools in this paper.

Literature [18] defined multimedia technology, which refers to the technology that uses image, sound, animation, and other visual materials to present in an intuitive form. When used in teaching, it plays a very important role in improving classroom teaching efficiency and enriching students' cognition. The use of multimedia technology in classroom teaching also proves the development of modern education. Based on the in-depth explanation of the main body of the music classroom, literature [19] discussed the application of multimedia technology in the classroom, and it is directly linked to the teaching effect. Moreover, it believes that multimedia technology should be widely popularized in music classrooms, which is of great significance for improving the efficiency of music classroom teaching. Literature [20] believed that, under the requirements of today's new curriculum reform, the compulsory education stage of music teaching has undergone profound changes in all aspects, especially for the cultivation of students' music literacy, which is the top priority and purpose of the reforms over the years. Literature [21] discussed the role of audiovisual artistic effects in music teaching from the perspective of multimedia technology applications. On this basis, it analyzed this issue in combination with corresponding teaching cases and believed that the effect of multimedia audiovisual art has a very positive significance for the efficiency of music classroom teaching.

3. Feature Extraction Technology for Music Frame Recognition

3.1. Popular Music Frame Feature Recognition. Each piece of music frame signal contains very rich personality characteristics of the singer, in addition to the short-term average energy, amplitude, and other time-domain characteristics. And there are many frequency domain parameters that can express various characteristics. If the characteristics of a piece of music frame signal are expressed more accurately, it is necessary to process the piece of music frame signal. Then the analysis of the signal to be measured must be completed after finding the starting end of the signal. The features extracted from unprocessed signals must be valuable characteristic coefficients, eliminating redundant information that is meaningless for music frame recognition, thereby weakening the amount of data to be analyzed in the subsequent recognition process.

Predictive coding usually uses a set of filters $H(z)$ to simulate channel characteristics. The specific algorithm of $H(z)$ is as shown in the following formula:

$$H(z) = \frac{X(z)}{E(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_{k^z} - k} = \frac{1}{A(z)}. \tag{1a}$$

Here, P is the order of LPC.

$$A(z) = 1 - \sum_{k=1}^{p} a_{k^z} - k. \tag{1b}$$

From formula (1), $x[n]$ can be calculated, as in the following formula:

$$x[n] = \sum_{k=1}^{p} a_k x[n-k] + e[n]. \tag{2}$$

When estimating the current sample size, the previous P samples are used for linear estimation. Therefore, this algorithm is called linear predictive coding. The specific algorithm is described in

$$\widehat{x}[n] = \sum_{k=1}^{p} a_k x[n-k]. \tag{3}$$

The error value $e[n]$ of linear predictive coding is shown in

$$e[n] = x[n] - \widehat{x}[n] - \sum_{k=1}^{p} a_k x[n-k]. \tag{4}$$

The algorithm calculates the logarithm on both sides of $H(z)$ and expands the Fourier series of z^{-1} to obtain the following formula:

$$\ln\left(\frac{1}{1 - \sum_{k=1}^{p} a_{k^z} - k}\right) = \sum_{n=1}^{\infty} c_{lp}(n) - n. \tag{5a}$$

Subsequently, the two sides of the equation are differentiated with respect to z^{-1}, and then we get

$$\frac{\sum_{k=1}^{p} k a_{k^z} - (i-1)}{1 - \sum_{k=1}^{p} a_{k^z} - k} = \sum_{n=1}^{\infty} n c_{lp}(n) z^{-(n-1)}. \tag{5b}$$

The equation is finally reduced to

$$\sum_{k=1}^{p} k a_{k^z} - (i-1) = \left(1 - \sum_{k=1}^{p} a_{k^z} - k\right) \sum_{n=1}^{\infty} n c_{lp}(n) z^{-(n-1)}. \tag{5c}$$

By modifying the exponents on both sides of the equation, the relationship between $c_{lp}(n)$ and $a_k\,(k = 1, 2, \ldots, p)$ can be obtained:

$$\begin{cases} c(1) = a_1, \\ c(n) = a_n + \sum_{i=1}^{n-1}\left(1 - \frac{k}{n}\right) a_k c(n-k), & 1 < n \leq p, \\ c(n) = \sum_{i=1}^{p}\left(1 - \frac{k}{n}\right) a_k c(n-k), & n > p. \end{cases} \tag{6}$$

It can be seen from (6) that the greater the order of the linear cepstrum, the greater the amount of information that can be preserved. In practical applications, since the LPC coefficients are linear signals at all frequencies, they are not consistent with the auditory characteristics of the human ear. This feature will affect the recognition ability of the music frame recognition system. In addition, the LPC coefficients contain a lot of high-frequency noise, which will also lead to a reduction in the discrimination ability of the music frame recognition system.

An amount of experimental data show that when the music frame signal does not exceed 1000 Hz, the perceptual performance of the human ear maintains a stable linear relationship with the frequency of the signal. However, when the frequency exceeds 1000 Hz, the relationship between the perception performance and the frequency becomes linear. The mel cepstrum coefficients are solved under the condition of using the filter bank. The width of this type of filter bank on the mel frequency scale is equal, and the MFCC spectrum is converted to coordinates based on the mel frequency. And its transformation formula is shown in (7a) or (7b):

$$F_{\mathrm{mel}} = 3322.23 \lg (1 + 0.001) f_{HZ}, \tag{7a}$$

$$F_{\mathrm{mel}} = 2595 \lg \frac{(1 + f_{HZ})}{700}. \tag{7b}$$

Here, F_{mel} represents the frequency in the mel spectrum, and f_{HZ} represents the frequency based on the ordinary spectrum. The specific process of solving the MFCC process is shown in Figure 1.

3.1.1. The Difference of MFCC.

As shown in the figure, the specific steps of MFCC include preemphasis, endpoint detection, framing, windowing, Fast Fourier Transformation (FFT), mel frequency filtering, Discrete Cosine Transform (DCT), and other key operations.

MFCC can only express the static characteristics of the sound signal. However, it is difficult to accurately extract the characteristics of singers using only one feature parameter. Only by combining dynamic and static features, it can achieve reliable performance requirements. In order to improve the recognition performance of the system, the difference of MFCC (Delta MFCC) is introduced, and its parameter d is as follows in the calculation formula:

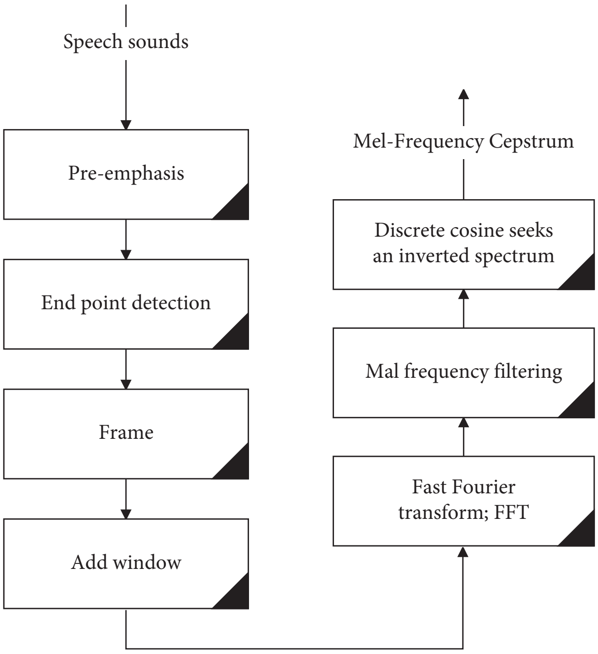

Figure 1: Flow chart of voice conversion into MFCC parameters.

$$d_i = \begin{cases} C_{i+1} - C_i, & t < K, \\ \dfrac{\sum_{k=1}^{K} k\left(C_{i+1} - C_i\right)}{\sqrt{2\sum_{k=1}^{K} k^2}}, & \text{other}, \\ C_{i+1} - C_i, & t \geq Q - K. \end{cases} \tag{8}$$

In (8), t represents the t-th time, C is the cepstral coefficient, Q is the order of the cepstral coefficient, and K represents the time difference.

Delta MFCC essentially weights and differentiates the feature parameters extracted by MFCC to further extract features. The voice features extracted twice by Delta MFCC can better express the continuous dynamic change law of the music frame signal, and it can also extract the difference multiple times to make the feature extraction effect better. In the music frame recognition system, generally the first-order difference is sufficient. The specific extraction process is shown in Figure 2.

From the difference flow chart in Figure 2 and the MFCC flow chart in Figure 1, it can be seen that adding the difference will actually add a step and increase the complexity of the calculation. From the application point of view, this paper studies a music frame recognition system based on mobile terminals. Taking into account the limitations of the mobile terminal itself, the difference of MFCC will increase the computational complexity and space complexity, and the effect may not be ideal in practical applications.

3.1.2. CMVN. The statistical characteristics of speech features are affected by the noise environment. Applying the normalization method to the speech recognition system can compensate for the impact of environmental noise mismatch and then improve the system's recognition rate. Most normalization methods are applied in the cepstrum frequency domain as the postprocessing of speech features. Among them, MFCC is widely used, and a variety of robust speech technologies are developed based on it. Its advantage is that it does not require any prior knowledge of noise environment and adaptive methods, it is easy to operate when implemented, and the effect is relatively ideal. One of the normalization methods is CMVN, that is, Cepstrum Mean Variance Normalization (CMVN). Its main purpose is to reduce the noise parameters in the voice channel, making the music frame signal more pure, and making feature extraction more efficient. The specific algorithm process of CMVN is as follows.

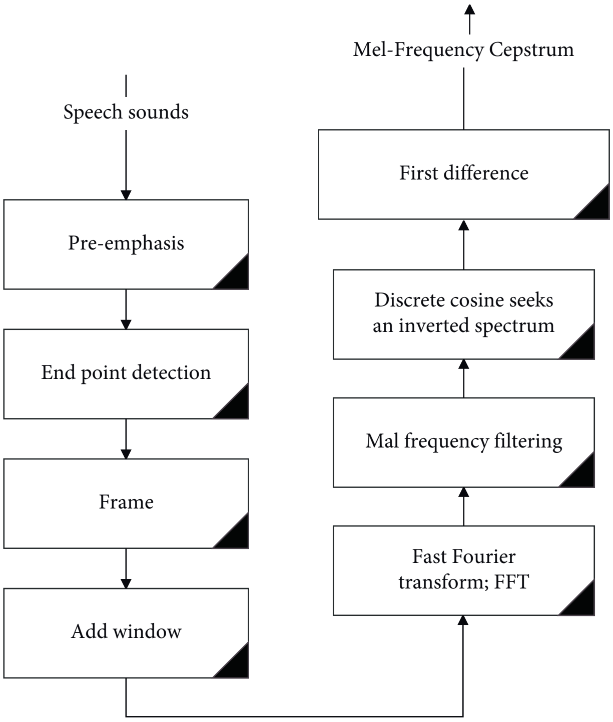

Figure 2: Differential parameter Delta MFCC extraction process.

First, we calculate its cepstral sequence $X(n)$. The Nth moment of $X(n)$ is defined as follows:

$$E\left[X^N(n)\right] = \frac{1}{T}\sum_{k=0}^{T-1} X^N(k). \tag{9}$$

In (9), $X(n)$ is the sequence of cepstral coefficients, and T is the sequence length.

Under normal circumstances, the cepstrum of the music frame signal is Gaussian distribution, then its odd-order moment value is 0, and the even-order moment is a certain constant. The characteristics of the cepstrum sequence normalized by the Nth moment are

$$E\left[X_{[N]}^N(n)\right] = \begin{cases} 0, & \text{N is odd number} \\ M_N, & \text{N is even number.} \end{cases} \tag{10}$$

According to (10), CMVN can be defined as

$$E\left[X_{[N]}^{N}(n)\right] = X_{1,2} = \frac{X_{[1]}(n)}{\sqrt{E\left[X_{[1]}^{2}(n)\right]}} = \frac{X(n) - E[X(n)]}{\sigma_x}. \tag{11}$$

Here, $X[L, N]$ is the corresponding sequence after the L and N moments of $X(n)$ are normalized at the same time.

3.2. Probabilistic Model of Music Frame Recognition. Building a probability model is a key step in music frame recognition. Therefore, the choice of probability model is very important to the effect of music frame recognition. Currently commonly used probability models are the following: based on dynamic time warping (DTW), vector quantization (VQ), HMM (hidden Markov model), and GMM. This section will introduce GMM and HMM.

GMM is a Gaussian mixture model. Although the voice distribution of singers is different, there is no strict curve law. However, almost all distributions can be approximated by using the mixture weight value of the Gaussian distribution, thereby obtaining a Gaussian mixture model.

$$\begin{aligned} P(X|\lambda_i) &= \sum_{i=1}^{M} w_i b_i(X), \\ \sum_{i=1}^{M} w_i &= 1. \end{aligned} \tag{12}$$

Here, X is a high-dimensional random speech feature vector; w, is the weighting coefficient of b and (X) corresponding components; and M is the number of components in the Gaussian mixture model. For $b_i(X)$ and w_i, they satisfy the following formula:

$$\begin{aligned} b_i(X) &= \frac{1}{(2\pi)^{(D/2)}\left|\sum_t\right|^{(1/2)}} \\ &\quad \cdot \exp\left\{-\frac{(X - u_t)^T \sum_t^{-1} (X - u_t)}{2}\right\}. \end{aligned} \tag{13}$$

We set λ_i as a parameter to describe a model, and we can get

$$\lambda_i = \left\{\omega_i, u_i, \sum_t\right\}, \quad i = 1, 2, 3, \ldots, M. \tag{14}$$

All parameter values should be classified according to a certain criterion, among which the maximum likelihood estimation algorithm is the most commonly used.

$$\log\, p(X|\lambda_i) = \sum_{t=1}^{T} \log\, p(x_t|\lambda_i). \tag{15}$$

The estimation of GMM parameter values is based on the maximum likelihood criterion (ML), which is realized by the EM iterative algorithm. At this time, the following iterative formulas for model weight, mean, and variance can be obtained:

$$\begin{aligned} w_i' &= \frac{1}{n}\sum_{j}^{n} p\left(i|x_j, \lambda_i\right), \\ u_i' &= \frac{\sum_j^n x_j p\left(i|x_j, \lambda_i\right)}{\sum_j^n p\left(i|x_j, \lambda_i\right)}, \\ \sigma_i'^2 &= \frac{\sum_j^n \left(x_j - u_i'\right)^2 p\left(i|x_j, \lambda_i\right)}{\sum_j^n p\left(i|x_j, \lambda_i\right)}. \end{aligned} \tag{16}$$

Here, $p(i|x_j, \lambda_i)$ is the posterior probability of the i-th mixed component.

HMM is different from GMM and is generally used to describe the statistical characteristics of the signal. When it is used to describe the statistical characteristics of a signal, two random processes are used. And these two random processes are interrelated. One of them is the hidden Markov chain, and the other is the random process of observing the vector, which is associated with each of the states of the Markov chain. After studying and observing the signal characteristics, the characteristics of the hidden Markov chain can be revealed. For continuous speech recognition, HMM is currently the best model and algorithm. And with the continuation of the research, its performance in all aspects is also constantly moving forward.

The HMM speech recognition process is shown in Figure 3.

As shown in Figure 3, HMM is divided into three problems: evaluation problem, identification problem, and training problem:

(1) We already know that the model $\lambda = f(A, B, \pi)$ and the output (the given observation sequence). When calculating the corresponding probability of generating Y, if there are several models to choose from, after solving this problem, the model that best matches the given observation sequence can be selected.

(2) We already know that the model $\lambda = f(A, B, \pi)$ and the output Y (how to estimate the input state X that is most likely to experience when this output is produced, that is, how to select the corresponding optimal observation sequence).

(3) If the model parameters are continuously revised by many outputs Y, the model parameters λ = fA, B) are optimized, so that the final model parameters are consistent with the output with the greatest probability.

The energy-zero-crossing rate two-stage fusion method can be achieved through the following steps.

3.2.1. The First Level of Judgment Is Energy Judgment. First, determine a threshold (in this paper, the empirical threshold) as the judgment standard, and compare the threshold T with the normalized metric T of the average energy of the music frame signal, and then determine whether the music frame signal is the singer's voice.

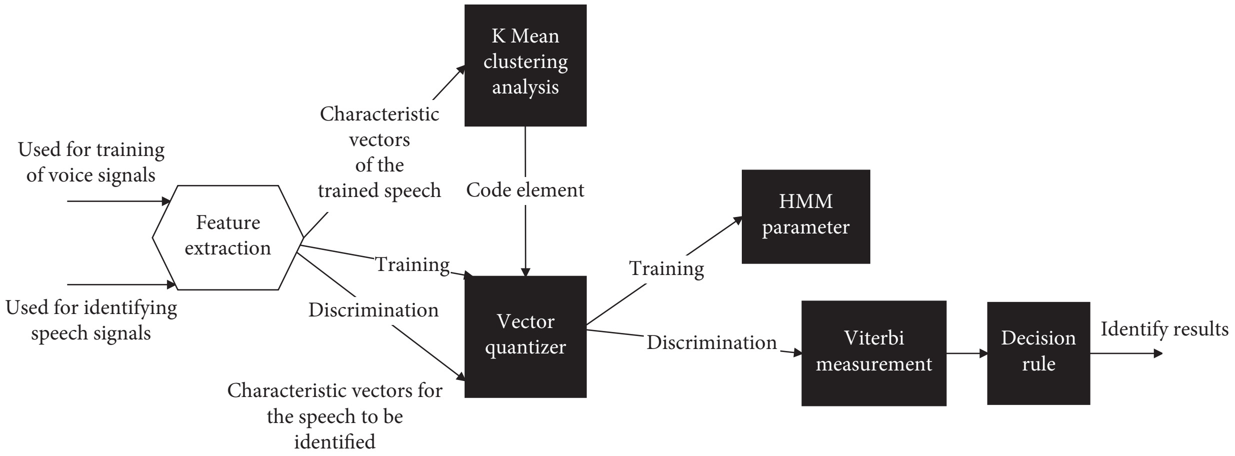
K Mean clustering analysis
Characteristic vectors of the trained speech
Used for training of voice signals
Code element
HMM parameter
Feature extraction
Training
Training
Used for identifying speech signals
Discrimination
Vector quantizer
Discrimination
Viterbi measurement
Decision rule
Identify results
Characteristic vectors for the speech to be identified

Figure 3: HMM music frame recognition process.

However, in general, the music frame signal can only maintain an approximately stable characteristic in a short period of time. Therefore, the music frame signal under short-term conditions can be treated as a steady signal. In order to obtain the short-term music frame signal, it is necessary to perform translation weighting on the original signal through a window of constant frame length, which is also the framing process of the music frame signal. In this paper, the Hamming window w(n) is used to divide the music frame signal, and its definition is in the following formula:

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), & (0 \le n \le N-1), \\ 0, & (n < 0 \text{ or } n < N). \end{cases} \tag{17}$$

Here, N is the length of the Hamming window. After framing processing, the short-term average energy of the music frame signal $x(n)$ can be obtained, that is, the average energy E of the speech frame, and its calculation formula is as follows:

$$E_n = \sum_{m=-\infty}^{+\infty} [x(m)w(n-m)]^2. \tag{18}$$

In (18), $x(m)$ is the sampled music frame signal. The normalized metric T of the short-term average energy of the music frame signal is defined as

$$T_2(i) = \frac{E_n(i)}{E_{n\_\max}}, \quad i = 1, 2, 3, \ldots, M. \tag{19}$$

Here, $A_{1\times M}$ is the maximum value of the average energy of all speech frames, i represents the i-th speech frame, and M represents the total number of speech frames.

$$E_{n\_\max} = \max E_n(i), \quad i = 1, 2, 3, \ldots, M. \tag{20}$$

If $T_2(i) > T_1$, the speech frame is judged to be the voice of the singer, and the coefficient a_i of the speech frame is assigned as "1." If $T_2(i) < T_1$, the speech frame is judged to be a nonsinger speech, and its coefficient $C = Z$ is defined. A piece of music frame signal has to go through M consecutive speech frame judgments, and an $I{\times}M$-order matrix M composed of "0″" and "1″" can be obtained. As the result of the first energy judgment, the definition of matrix M is as follows:

$$A_{1\times M} = [a_1 a_2 a_3 \ldots a_i \ldots a_M]. \tag{21}$$

3.2.2. The Second Level of Judgment Is Zero-Crossing Rate Judgment. Since the music frame signal is very unstable and can only remain stable within a short period of time, the zero-crossing rate judgment of the music frame signal is also involved in the short-term zero-crossing rate. The judgment rules are as follows.

The algorithm selects the threshold ZCR (in this paper, the empirical threshold 5 is selected). If the zero-crossing rate of the voice frame is lower than the threshold, it means that there is no singer's voice, and the coefficient b_i of the frame is assigned as "0." Conversely, if it is higher than the threshold, the coefficient b_i of the frame is assigned a value of "1." That is, the judgment result $B_{1\times M}$ of the second zero-crossing rate is obtained as the following formula:

$$B_{1\times M} = [b_1 b_2 b_3 \ldots b_i \ldots b_M]. \tag{22}$$

The zero-crossing rate formula is as the following formula:

$$Z_n(i) = \frac{1}{2} \sum_{m=-\infty}^{+\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(n-m), \quad i = 1, 2, 3, \ldots M. \tag{23}$$

In formula (23), sgn[] is a symbolic function, expressed as the following formula:

$$\operatorname{sgn}[x(n)] = \begin{cases} 1, & (x > 0), \\ -1, & (x < 0). \end{cases} \tag{24}$$

4. Judgment Fusion

Step 1. It first takes the union of the two judgment results $A_{1\times M}$ and $B_{1\times M}$ to obtain the preliminary fusion result $\widehat{R}_{1\times M}$, as shown in

$$\widehat{R}_{1\times M} = A_{1\times M} \cup B_{1\times M} = [r_1, r_2, r_3, \ldots, r_i, \ldots, r_M]. \tag{25}$$

Here, r_i is the speech frame coefficient after fusion.

Step 2. If there is such a speech frame $x_i(n)$: in the first energy judgment, it is judged to be the speech frame of the nonsinger's speech (that is, $a_i = 0$), and in the second zero-crossing rate judgment, it is judged to be the speech frame of the singer's speech voice frame (that is, $b_i = 1$). Then the coefficient r after the fusion of the two is cleared to zero. In other words, if $a_i = 0$ and $b_i = 1$, then we set $r_i = 0$. In this way, the detection result $R_{1\times M}$ can be obtained, which is also the final result of the two-stage fusion endpoint detection method based on energy and zero-crossing rate. The specific process of this improved algorithm is shown in Figure 4.

According to the algorithm flow in Figure 4, an example is given to illustrate the two-stage fusion endpoint detection method based on energy and zero-crossing rate. For example, if the energy judgment result of a certain segment of speech is

$$A_{1\times 12} = [a_1 a_2 a_3 \ldots a_{12}] - [\,000 \;\; 110 \;\; 001 \;\; 100\,]. \tag{26}$$

And, the result of the zero-crossing rate judgment is

$$B_{1\times 12} = [b_1 b_2 b_3 \ldots b_{12}] = [\,001 \;\; 100 \;\; 100 \;\; 001\,]. \tag{27}$$

Then, its preliminary fusion results are

$$\begin{aligned} R_{1\times 12} &= A_{1\times 12} \cup B_{1\times 12} = [r_1 r_2 r_3 \ldots r_{12}] \\ &= [\,001 \;\; 110 \;\; 101 \;\; 101\,]. \end{aligned} \tag{28}$$

Since $a_3 = 0$ and $b_3 = 1$, $a_7 = 0$ and $b_7 = 1$, $a_{12} = 0$ and $b_{12} = 1$, r_3, r_7, r_{12} are all assigned the values "0." The final endpoint detection result is

$$R_{1\times 12} = [\,000 \;\; 110 \;\; 001 \;\; 100\,]. \tag{29}$$

The energy-zero-crossing rate two-stage fusion detection algorithm proposed in this section is applied to the actual music frame signal endpoint detection, and the detection effect is shown in Figure 5.

In the actual detection effect experiment shown in Figure 5, the upper part of the music frame is the input unprocessed voice, and the lower part is the voice after endpoint detection. It can be seen that the endpoint detection accurately cuts out the effective voice segment. Among them, the tested singers were asked to speak 10 two-syllable words within 20 seconds.

5. Evaluation of Multimedia Popular Music Teaching Effect Integrating Semifinished Teaching Strategies

The "semifinished product processing" strategy is an advanced teaching concept and method that has become more popular in recent years. It refers to providing teachers and students with partially completed teaching works, that is, "semifinished products." It can be understood as a kind of technology "retaining" the complete teaching results (finished products), simulating the real problem solving environment, supplementing these blanks which is "reprocessing" and teaching in the process of forming the "finished products." Using the "semifinished product processing" teaching method, neither teaching nor practice will destroy the authenticity and integrity of the overall work. Moreover, it simplifies the process of reoperation, solves the problem of the difference in the progress of the students' operation, and ensures the continuity and integrity of the teaching content and the student's operation technology. At the same time, it can enable students to master the essentials of each operation link in a short time, improve the efficiency of teachers' regular teaching explanations, students' understanding of concepts, and practical exercises, and optimize the learning situation and training environment.

This article uses B/S framework, SQL server database for development, and SOA service framework as the overall structure of the system. Taking into account the overall integration of each module and the system, a four-tier architecture is used to design the system framework, namely, the data storage layer, the data access layer, the interface service layer, and the business function layer. Among them, the data storage layer saves the basic data and business data of the entire systems and uses the SQL server database to manage the data. The data access layer provides data interface services for the system through SQL language. The interface service layer is written using the J2EE service framework, provides Weblogic services, and provides interfaces for the various functions of the system to retrieve and download data resources. The business application layer provides all system functions used by users, including system management, online Q&A, online examination, course resource management, and user login. The interaction with the user is realized through the WEB interface, as shown in Figure 6.

The audio processing system includes two parts: a feature extraction module and a feature processing module. The feature extraction module includes a pitch frequency recognizer and a pitch acquirer, as shown in Figure 7. The built-in pitch frequency recognition methods of the pitch frequency recognizer include cepstrum method, harmonic peak method, cyclic direct method, wavelet transform method, and parallel processing method. According to the method of obtaining the pitch selected by the teacher in the teacher guidance system, the pitch frequency data feature is extracted.

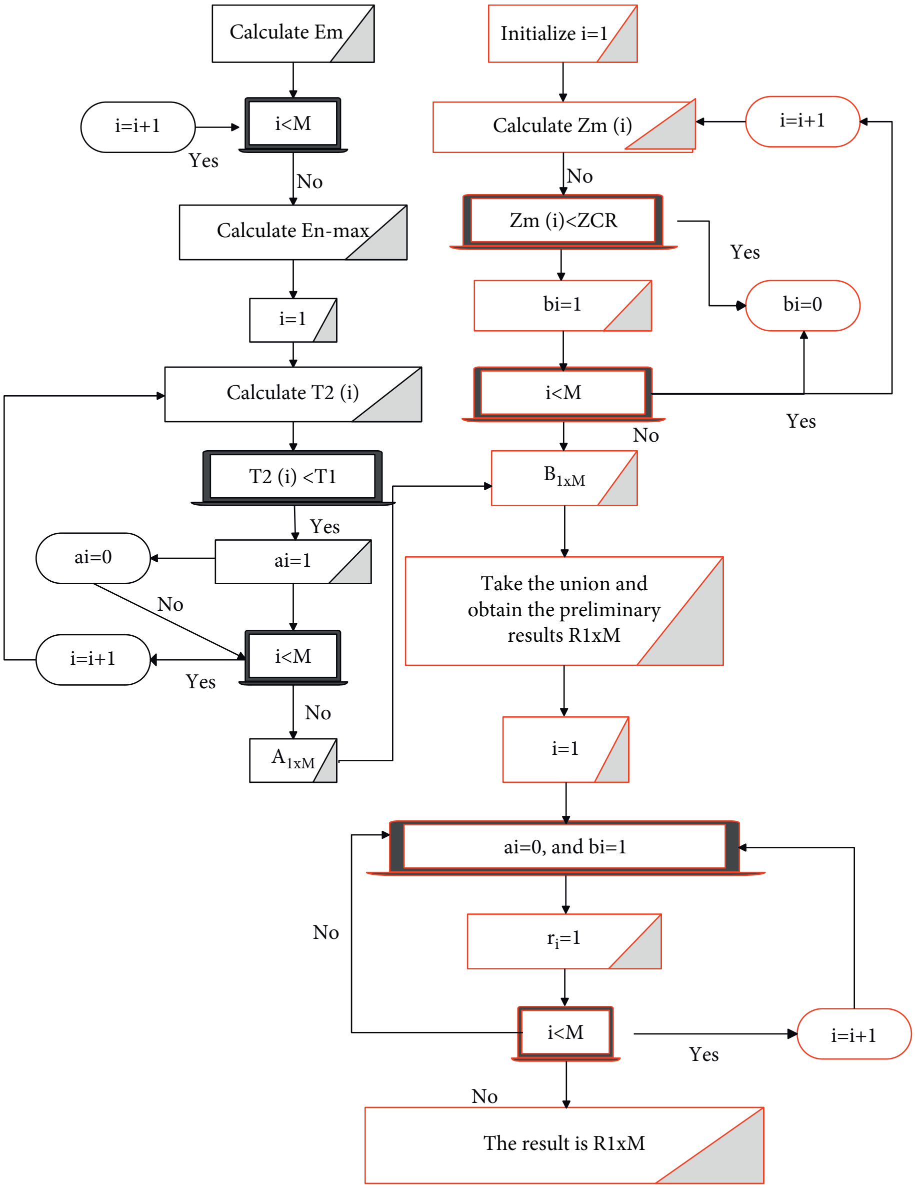

Figure 4: Flow chart of two-stage fusion method of energy-zero-crossing rate.

The multimedia pop music teaching system integrated with the semifinished product teaching strategy mainly uses the B/S three-tier system structure, as shown in Figure 8.

After constructing a multimedia pop music teaching system that integrates semifinished product teaching strategies, the performance of the system is verified. This article mainly examines from two aspects of music frame feature recognition and music teaching effect and conducts research through simulation experiments. Construct a simulation system model through Matlab and perform system performance evaluation and analysis combined with expert evaluation methods. Compare the method in this article with literature [11]; the results are shown in Tables 1 and 2 and Figures 9 and 10.

From the above research, it can be seen that the multimedia pop music teaching system integrating semifinished product teaching strategies proposed in this paper can play an important role in music teaching and effectively improve the effect of music teaching.

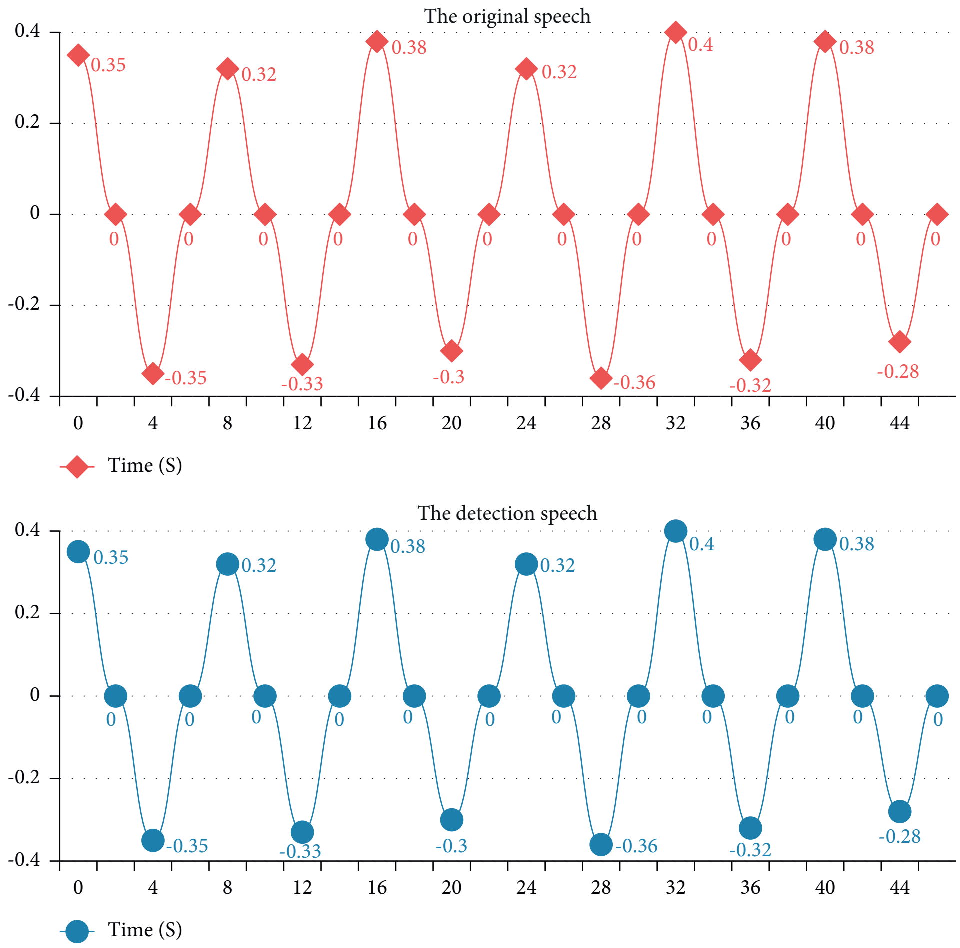

Figure 5: The effect of the improved algorithm in the actual endpoint detection. (a) The original speech. (b) The detection speech.

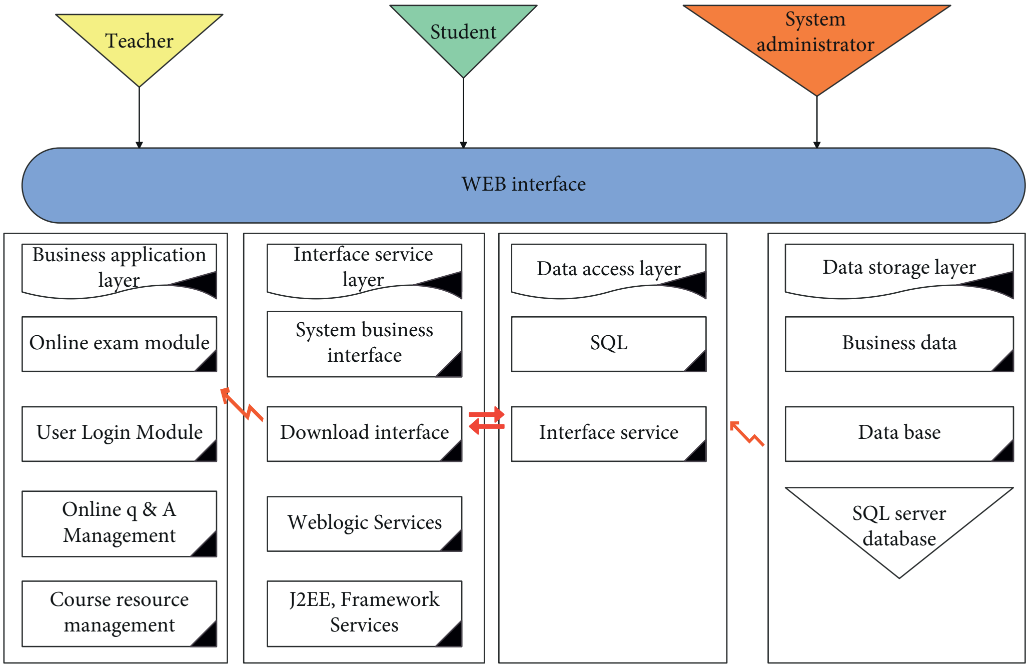

Figure 6: Multimedia popular music teaching system incorporating semifinished teaching strategies.

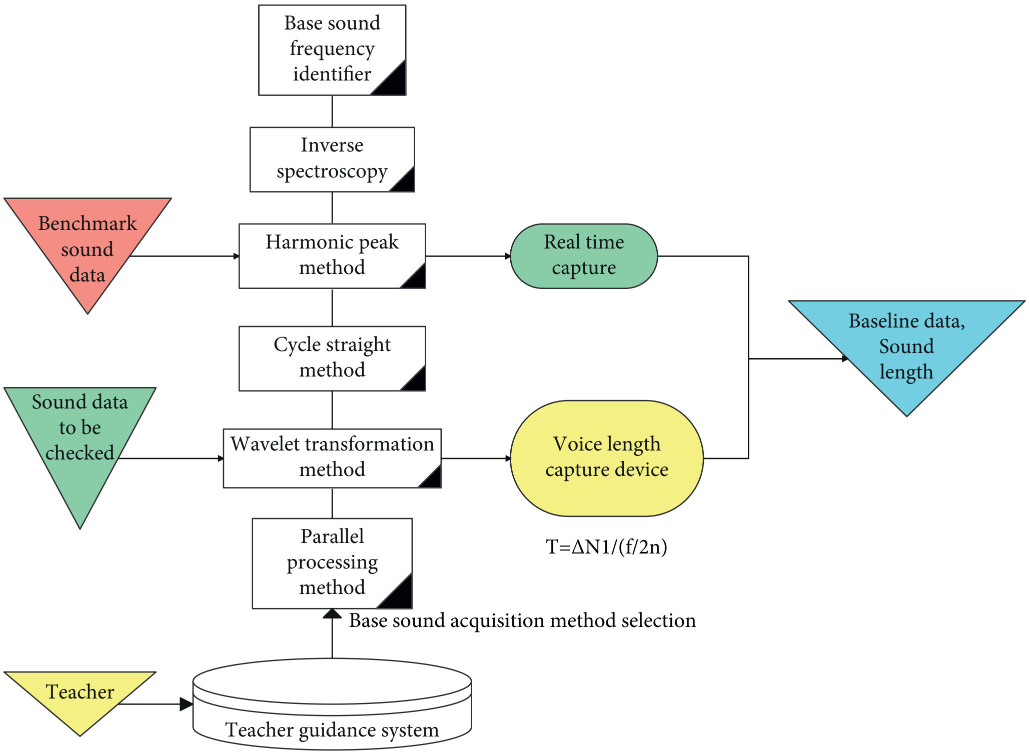
Base sound frequency identifier
Inverse spectroscopy
Benchmark sound data
Harmonic peak method
Real time capture
Baseline data, Sound length
Cycle straight method
Sound data to be checked
Wavelet transformation method
Voice length capture device
Parallel processing method
T=ΔN1/(f/2n)
Base sound acquisition method selection
Teacher
Teacher guidance system

Figure 7: Feature extraction module.

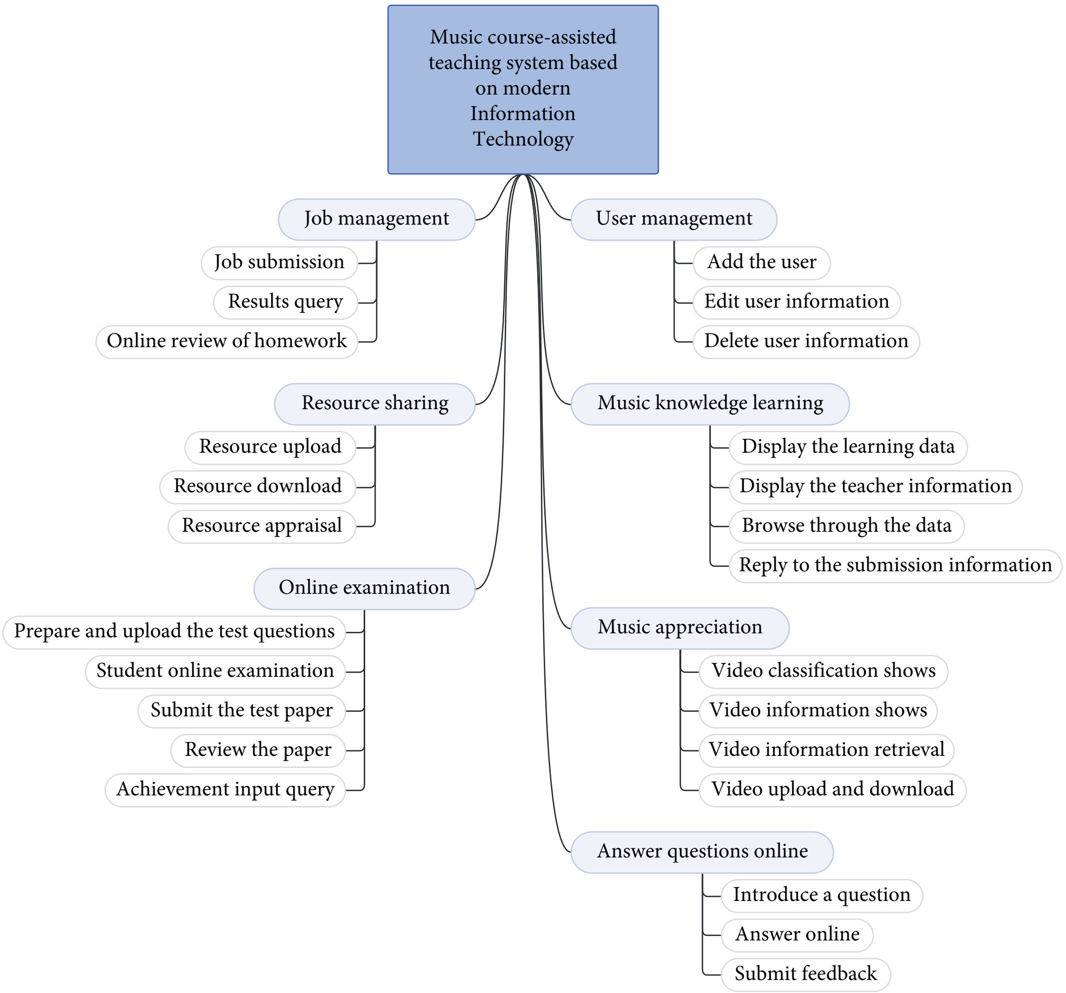

Figure 8: Function module diagram of the music teaching system.

Table 1: Music frame feature recognition effect comparison.

Num	Method of this article	Method of literature [11]
1	91.58	82.15
2	86.54	77.27
3	89.21	69.97
4	83.23	60.29
5	86.16	68.70
6	88.74	64.42
7	93.90	79.65
8	90.82	80.06
9	86.40	70.59
10	88.37	68.71
11	91.10	73.17
12	91.48	67.85
13	93.61	83.77
14	88.57	67.71
15	93.80	83.12
16	87.95	74.03
17	83.19	62.06
18	93.24	70.30
19	90.47	73.58
20	84.79	75.61

Table 2: Music teaching effect comparison.

Num	Method of this article	Method of literature [11]
1	84.86	73.37
2	88.25	68.93
3	76.64	56.22
4	84.80	72.30
5	78.39	61.35
6	80.86	61.61
7	77.37	67.67
8	82.81	68.67
9	82.62	69.21
10	76.82	57.74
11	76.05	62.05
12	78.81	57.22
13	89.86	75.50
14	76.20	64.75
15	75.52	66.74
16	79.78	66.86
17	75.54	61.08
18	78.43	69.58
19	85.14	63.51
20	80.45	64.34

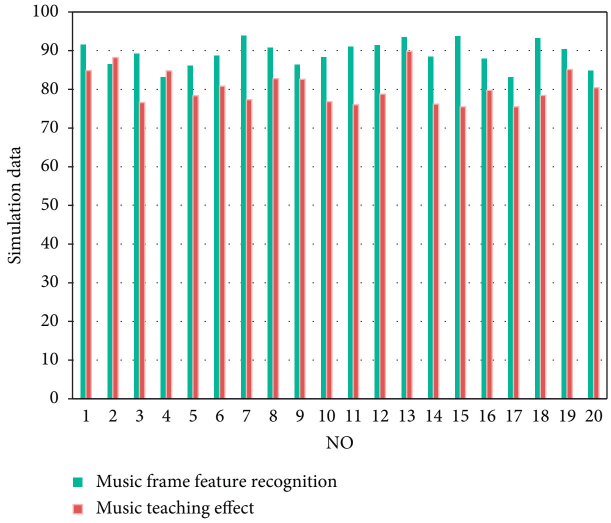

FIGURE 9: Music frame feature recognition effect comparison statistics.

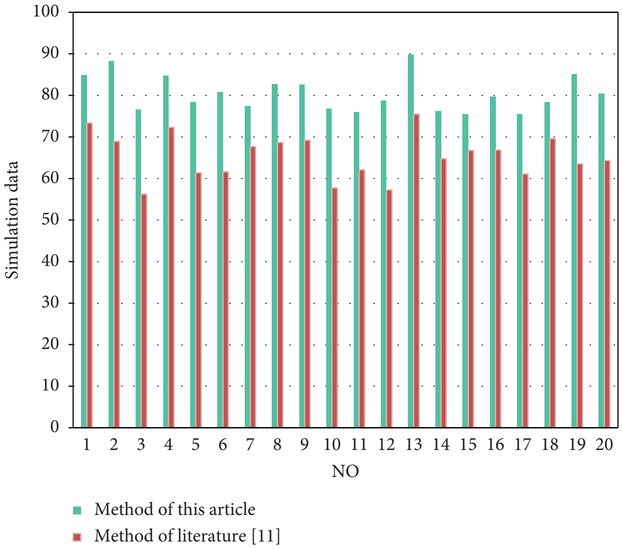

FIGURE 10: Music teaching effect comparative statistics.

6. Conclusion

From the perspective of development momentum, the education of popular music will become an indispensable part of the education system of various music colleges, and the teaching of sight singing and ear training as the basis of various disciplines is obligatory to incorporate popular music elements into the teaching. Music learning is different from other subjects. Learners need to be effectively guided, and they need to use more flexible teaching methods to let students understand the abstract music teaching content. In today's social and cultural context, popular music continues to be popularized, and there are more and more people who love popular music, and the trend is gradually rising. As the basic course of music learning, sight singing and ear training should incorporate popular music into the teaching and then expand the teaching content and teaching space to enhance students' interest in learning. Based on this, this paper integrates the teaching strategies of semifinished products to construct a multimedia pop music teaching system and evaluates the teaching effect of the system, so as to provide a theoretical reference for the follow-up pop music teaching.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This study was sponsored by Xinyang College.

References

[1] J. Chin, V. Callaghan, and S. B. Allouch, "The internet-of-things: reflections on the past, present and future from a user-centered and smart environment perspective," *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 1, pp. 45–69, 2019.

[2] G. Bedi, G. K. Venayagamoorthy, R. Singh, R. R. Brooks, and K.-C. Wang, "Review of internet of things (IoT) in electric power and energy systems," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 847–870, 2018.

[3] I. Bisio, A. Delfino, A. Grattarola, F. Lavagetto, and A. Sciarrone, "Ultrasounds-based context sensing method and applications over the internet of things," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3876–3890, 2018.

[4] A. Chamberlain, M. Bødker, A. Hazzard et al., "Audio technology and mobile human computer interaction," *International Journal of Mobile Human Computer Interaction*, vol. 9, no. 4, pp. 25–40, 2017.

[5] D. B. Ç Kiliç, "Pre-service music teachers' metaphorical perceptions of the concept of a music teaching program," *Journal of Education and Learning*, vol. 6, no. 3, pp. 273–286, 2017.

[6] D. L. Hoffman and T. P. Novak, "Consumer and object experience in the internet of things: an assemblage theory approach," *Journal of Consumer Research*, vol. 44, no. 6, pp. 1178–1204, 2018.

[7] B. Jia, L. Hao, C. Zhang, H. Zhao, and M. Khan, "An IoT service aggregation method based on dynamic planning for QoE restraints," *Mobile Networks and Applications*, vol. 24, no. 1, pp. 25–33, 2019.

[8] J. Waldron, R. Mantie, H. Partti, and E. S. Tobias, "A brave new world: theory to practice in participatory culture and music learning and teaching," *Music Education Research*, vol. 20, no. 3, pp. 289–304, 2018.

[9] J. Zhang and D. Tao, "Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7789–7817, 2020.

[10] E. Gun, "The opinions of the preservice music teachers regarding the teaching of orchestra and chamber music courses during distance education process," *Cypriot Journal of Educational Sciences*, vol. 16, no. 3, pp. 1088–1096, 2021.

[11] G. Muhammad, S. M. M. Rahman, A. Alelaiwi, and A. Alamri, "Smart health solution integrating IoT and cloud: a case study of voice pathology monitoring," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 69–73, 2017.

[12] X. Shengmin, "Analysis on the innovative strategy of national music teaching in colleges from the perspective of visual communication," *Studies in Sociology of Science*, vol. 7, no. 6, pp. 52–55, 2017.

[13] Z. Lian, "Research on aesthetic education in instrumental music teaching," *Journal of Literature and Art Studies*, vol. 10, no. 5, pp. 435–439, 2020.

[14] S. K. Kim, N. Sahu, and M. Preda, "Beginning of a new standard: internet of media things," *KSII Transactions on internet and information systems*, vol. 11, no. 11, pp. 5182–5199, 2017.

[15] A. Kaplan and M. Haenlein, "Siri, Siri, in my hand: who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence," *Business Horizons*, vol. 62, no. 1, pp. 15–25, 2019.

[16] F. L. Reyes, "A community music approach to popular music teaching in formal music education," *The Canadian Music Educator*, vol. 59, no. 1, pp. 23–29, 2017.

[17] V. K. Jones, "Voice-activated change: marketing in the age of artificial intelligence and virtual assistants," *Journal of Brand Strategy*, vol. 7, no. 3, pp. 233–245, 2018.

[18] P. S. Aithal and S. Aithal, "Management of ICCT underlying technologies used for digital service innovation," *International Journal of Management, Technology, and Social Sciences*, vol. 4, no. 2, pp. 110–136, 2019.

[19] C. Johnson, "Teaching music online: changing pedagogical approach when moving to the online environment," *London Review of Education*, vol. 15, no. 3, pp. 439–456, 2017.

[20] P. L. Lin, "Trends of internationalization in China's higher education: opportunities and challenges," *US-China Education Review B*, vol. 9, no. 1, pp. 1–12, 2019.

[21] S. Y. Hong and Y. H. Hwang, "Design and implementation for iort based remote control robot using block-based programming," *Issues in Information Systems*, vol. 21, no. 4, pp. 317–330, 2020.