Scholar Publishing Group
International Journal of Educational Innovation and Science
https://doi.org/10.38007/IJEIS.2023.040107
ISSN 2789-7184 Vol. 4, Issue 1: 82-98

Design and Implementation of Personalized Teaching System for Ethnic Vocal Music Learning Resources Based on Computer Vision

Huiwen Jia*

Lingsheng Technology Co., LTD, Beijing, China

mup05039@126.com

**corresponding author*

Keywords: Ethnic Vocal Music, Personalized Teaching System, Computer Vision, Learning Resources

Abstract: National vocal music is a traditional Chinese music genre and has a great influence in China. However, the current teaching of national vocal music is still conducted in the traditional small-class teacher-apprentice relationship, which hinders the sustainable development of national vocal music. Therefore, this paper aimed to use computer vision technology to build a national vocal music learning resource library and design a personalized teaching system for learning resources. For computer vision technology, this paper used image segmentation and video tracking technology to design an image recognition algorithm for hand and face, which provided help for students in the classroom. For the design of personalized teaching system, this paper introduced the database of ethnic vocal music learning resources and the steps of using vocal music learning resources in detail, and designed the B/S structure of the system in detail. From the test results of the system in this paper, it can be seen that the test success rate of the five functional modules of the system was above 99%, and the response time and delay of the system were within the acceptable range, which showed that the system designed in this paper could provide certain help for the teaching of national vocal music.

1. Introduction

There are two definitions of national vocal music. The narrow national vocal music focuses on traditional arts, such as folk art, rap, etc. The content is relatively limited and the form is very fixed. The broad national vocal music refers to an art form that is vigorously promoted and promoted in

Chinese colleges and universities based on the traditional Chinese artistic style and characteristics, and by borrowing from European vocal music theories and singing methods. Therefore, compared with other countries in the world, China's vocal music is richer and more diverse in both content and expression. National vocal music has mainly experienced two stages of development: traditional national vocal music and modern national vocal music. After entering the 21st century, more and more people began to pay attention to national vocal music, and national vocal music received more attention. National vocal music in colleges and universities has also received great attention. During this period, both the vocal music teachers in colleges and the form of vocal music teaching in colleges have been greatly developed, and many people began to specialize in the teaching of ethnic vocal music and established professional research centers. However, the current national vocal music teaching is still too traditional, and the masses' understanding of national vocal music is limited to a few famous national vocal artists. Therefore, the design and research of the personalized teaching system of ethnic vocal music learning resources based on computer vision is particularly important.

Ethnic vocal music is an important type of music in China. It has precipitated and inherited the spirit of China for thousands of years. There are also many related studies on ethnic vocal music. Combined with his understanding of intangible cultural heritage and his thinking on the promotion and inheritance of ethnic music from the perspective of intangible cultural heritage, Lu K further discussed the significance and strategy of promoting and inheriting ethnic music with vocal music teaching as the core [1]. Innovative integration is the development trend of education reform in recent years. Zhou Y tried to promote the high integration of aesthetic education and moral education from the establishment of a music studio, forming a "three-in-one" education and education integration mechanism [2]. Harmony is an essential element of music, speech, and animal vocalization. Lei F studied how the auditory system extracts the harmonic structures embedded in complex sounds and used them to form a coherent single entity [3]. The purpose of Liu X was to determine the details of the vocal narrative in terms of genre [4]. Kang H discussed the selection method of vocal music works in college vocal music teaching, and on the basis of previous research on vocal music teaching, Kang H conducted in-depth research on effective teaching methods in college vocal music teaching [5]. However, their research on ethnic vocal music pays more attention to the teaching theories of famous ethnic vocal musicians, and does not combine the arts of the masses.

In the Internet era, everything is inseparable from computer and network technology, and there are many researches on personalized teaching systems. Based on computer vision technology, Li Y had studied a new type of personalized college English teaching system based on information technology. The designed system aimed to improve the teaching quality of college English and stimulate students' enthusiasm for English learning [6]. Zang R proposed a new method for talent cultivation, and combined computer technology to analyze and allocate training programs in a unified way [7]. Marinagi C had designed a personalised teaching system for the learning of students with disabilities, a web-based tool tailored for students with disabilities. The experimental tools include the application of traditional learning management systems (LMS) supporting multilingual term dictionaries, multimedia management and social networking [8]. Zhao put forward new insights on the algorithm of the current personalized teaching system, and proposed a CFR algorithm combining user attributes and interest topic similarity [9]. Tang Y exemplified an important component of parallel intelligent education—artificial education systems that provide personalized learning in a narrative game environment [10]. However, at present, few scholars combine the computer system with the teaching of ethnic vocal music, and there are only few innovations in the teaching of ethnic vocal music.

In today's era of rapid development of information technology, the field of education and

teaching should keep pace with the development of information technology and boldly use new technologies to create an information-based modern educational environment, so as to push the quality of modern education and teaching to a higher level. In view of the current situation of minority music teaching and few learning resources, this paper designed a personalized teaching system for ethnic music based on computer vision technology.

2. National Vocal Music and Computer Vision Technology

2.1. National Vocal Music Teaching

With the development of the economy, people's pursuit of the spiritual world has gradually increased. As an art form, national vocal music has gradually developed into a professional art course and has begun to develop in Chinese universities. At first, the main people engaged in the teaching of modern national vocal music were the vocal music workers who were retained from Europe, they learned and taught while training, and trained a large number of vocal music talents for China [11]. The rich ethnic music types are shown in Figure 1.

Figure 1. Rich ethnic vocal expressions

National vocal music is an art form that shows the character and national characteristics of Chinese sons and daughters with Chinese culture as the background and national language as the basis. The characteristics of national vocal music are shown in Figure 2.

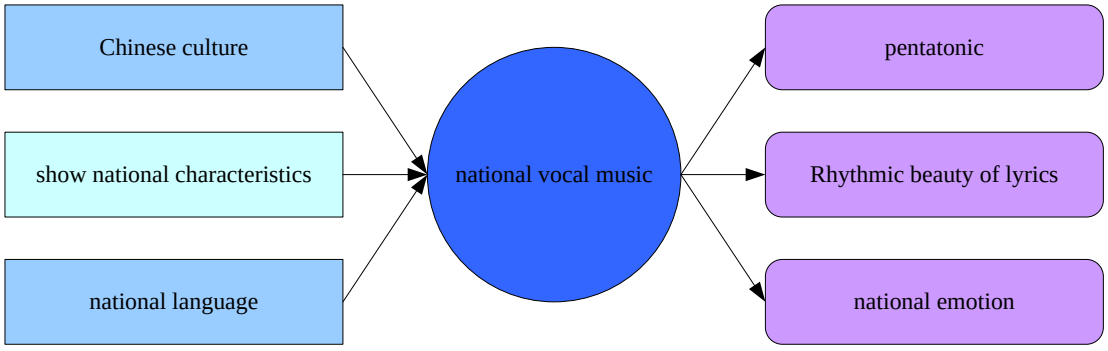


Figure 2. Characteristics of national vocal music

First, in terms of the creation of vocal music works, influenced by traditional vocal music, modern national vocal music inherited and developed the pentatonic tones (Gong, Shang, Jiao, Zheng, Yu), and created the six-tone and seven-tone scales on this basis. Second, in terms of the language of vocal works, Chinese national vocal music also pays great attention to the rhythmic beauty of the lyrics, and many musical works have a catchy, smooth and light feeling. Third, in terms of emotional expression of vocal works, only through the performance of singers can one feel the style of the emotions of their works. Through the performance of vocal singers, the emotions and emotions conveyed by songs can also be felt. Vocal singers in each country are used to interpreting songs in their own way of expressing emotions, so that the audience can understand and listen to them [12].

2.2. Application of Computer Vision Technology in National Vocal Music

As a comprehensive stage singing art, the successful performance of vocal music is not only the musical language, but also the physical performance. As shown in Figure 3, when students sing their works, they incorporate appropriate body language to effectively enhance the appeal of the works.

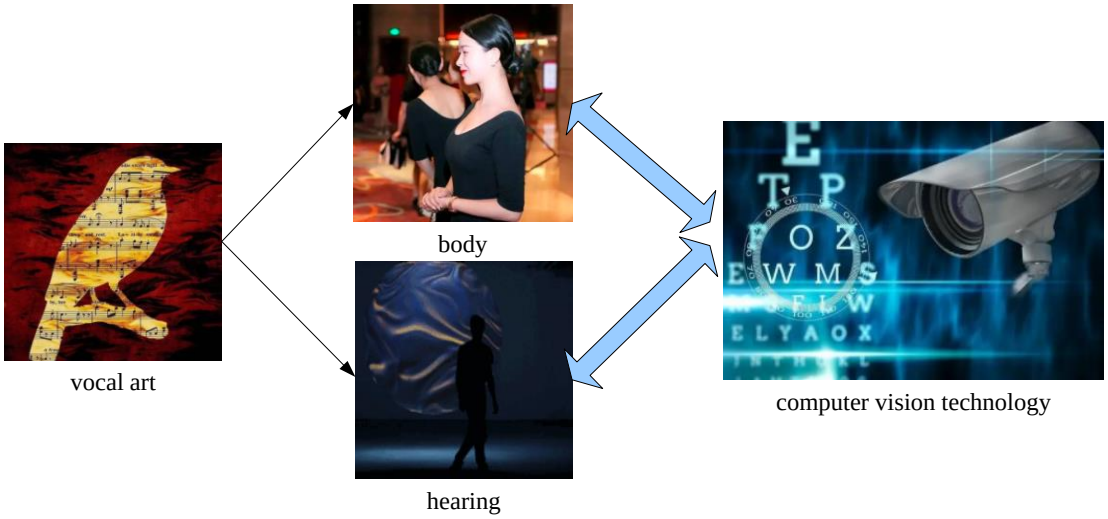


Figure 3. Application of computer vision technology in vocal art

The expression of "shape", the vocal art is closely linked with the visible body expression and the auditory art. The expression of body language is more intuitive, and it can reflect the inner activities of the singer. The expression of rich body language is not achieved overnight, it requires a long period of accumulation and learning, and it also requires learners to experience and think with their hearts. The expression of body language must first be conceived in advance according to the content of the vocal music [13].

2.3. Face and Hand Recognition Algorithms

At present, the research on interaction technology includes a variety of methods, including face interaction, gesture interaction, voice interaction, and facial expression interaction. In human daily life, gesture information has the advantages of intuition and simplicity, and has become a research hotspot. Hand information data collection methods are divided into two methods: based on data gloves and based on computer vision. Since identification based on data gloves requires the wearing of complex data gloves and tracking equipment, and the use cost is high, this method is not adopted

in this paper. With the development of image processing technology and the improvement of computer processing speed, hand recognition technology based on computer vision is more and more favored by people, and has attracted the attention of researchers [14].

(1) Image Preprocessing

Computer vision-based hand recognition is based on image processing, and the quality of the image affects the extraction of hand features. Since the image captured by the camera will inevitably introduce noise interference, it is necessary to perform noise reduction processing on the image before hand image segmentation. In the subsequent processing of hand image segmentation, in order to avoid the interference of non-target areas on feature point extraction, non-target areas need to be pruned. In this paper, processing techniques such as image filtering, image binarization and image morphological processing were used to process the images before and after segmentation of the hand image [15].

Mean filtering is a relatively simple linear filtering method, which uses the idea of neighborhood averaging. The algorithm will calculate the average value of the gray value of each point on the neighborhood of a certain pixel point of the image, and use it as the gray value of the point [16]. For example, it can define an image as $f(i, j)$ and (i, j) as the pixels on the image, then the expression corresponding to the mean filter is:

$$g(i, j) = \frac{1}{n} \sum_{i, j \in M} f(i, j) \tag{1}$$

Among them, M is the value range of image pixel points covered by the calculation template, and n is the number of pixels included in the neighborhood of point (i, j) covered by the calculation template operator.

Assuming that the size of the Gaussian filter template is $m \times n$, then the Gaussian filter template can be expressed as:

$$g(i, j) = \frac{1}{2\pi\sigma^2} e^{\frac{(t-m/2)^2+(j-n/2)^2}{2\sigma^2}} \tag{2}$$

Define the image as $f(x, y)$ and the two-dimensional template as M, then use the template operator to median filter the image $f(x, y)$, and the obtained image is:

$$g(x, y) = Med_{(x,y) \in N} f(x, y) \tag{3}$$

Among them, N is the value range of the image pixel position covered by the calculation template M, and Med represents the median operation of all pixel values in the template.

Image Morphological Processing

In the system designed in this paper, the following algorithms were used: erosion and dilation, opening operation, and closing operation [17].

The erosion algorithm mainly requires the minimum value of the area, and uses this value as the value of the specified pixel point. Assuming that a target area to be processed is A, and B is a structural element, convolve A and B to obtain a local minimum value. The corrosion of the structural element B to the target area A is recorded as $A \Theta B$, then the corrosion can be defined as:

$$A \Theta B = \{x, y | B_{xy} \subseteq A\} \tag{4}$$

The expansion algorithm mainly requires the maximum value of the area, and uses this value as the value of the specified pixel point. Assuming that a target area to be processed is A, and B is a

structural element, convolve A and B to obtain a local maximum. The erosion of the target area A by the structural element B is denoted as $A \oplus B$, then the expansion can be defined as:

$$A \oplus B = \{x, y | B_{xy} \cap A \neq \phi\} \tag{5}$$

In image morphological processing, the opening operation is a combined operation of erosion and dilation operations. The opening operation is to erode the image first, and then perform the dilation operation on the image. Use the structure element B to perform the opening operation on the target area A, which can be denoted as $A \circ B$, and the definition can be expressed as:

$$A \circ B = (A \Theta B) \oplus B \tag{6}$$

The closing operation is also a combined operation of the erosion and dilation operations, but different from the opening operation, the closing operation first performs the dilation operation on the image, and then performs the erosion operation on the image. Use the structure element B to perform the closing operation on the target area A, which can be denoted as $A \bullet B$, and the definition can be expressed as:

$$A \bullet B = (A \oplus B) \Theta B \tag{7}$$

(2) Hand Image Segmentation Based on Skin Color

Human gestures are one of the most natural ways for human beings to express their wishes and have rich semantic information. Human-computer interaction based on human gestures has high value in augmented reality interaction methods [18].

The display of color images is the expression of colors based on the color space. There are many kinds of color spaces, the most widely used RGB color space at present, in which R represents the red component, G represents the green component, and B represents the blue component. The RGB color space is suitable for real systems, and the color images of digital cameras are also based on the RGB space. However, the detection and segmentation of images is not suitable for RGB color space, because RGB color space cannot separate lighting factors, so it is greatly affected by lighting. In the analysis of the three components of the RGB color space, it can be seen that the clustering of the R, G, and B components is not very obvious, so it is necessary to convert the color space.

Since the detection of hand images is more accurate in the YCRCB color space, the system studied in this paper needs to convert the RGB color space to the YCRCB color space. The conversion of RGB color space to YCRCB color space can be expressed as:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \bullet \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{8}$$

Figure 4. Segmentation renderings

As shown in Figure 4, the hand image segmentation using the algorithm in this paper has a good effect, can better extract the outline of the hand, the arm part can also be more clearly highlighted, and the segmentation of the background can also be better presented.

Face Detection and Recognition Process

Adaboost technology is a commonly used classifier in face detection algorithms, and has a complete code in the OpenCV vision library, which is easier to retrieve, the method is as:

$$feature_j = \sum_{i \in (1,2)} w_i \operatorname{Re} ctSum(r_i) \tag{9}$$

The integral graph is a method that can be $\operatorname{Re} ctSum(r_i)$, and the operation efficiency is high. For any face image, the gray level of each pixel of the other image is i(x,y), then the value ii(x,y) of the pixel corresponding to the corresponding integral map can be obtained according to the corresponding intermediate variables x' and y' in i(x,y), and the calculation method is shown in Formula 10.

$$ii(x, y) = \sum_{x' \le x, y' \le y} i(x', y') \tag{10}$$

In order to solve the face shape S, it is assumed that there is an initial state S_0 for the face alignment algorithm. For the face shape S, its change process is determined with the shape variable ΔS, where ΔS is a very small variable. Therefore, the change process of the face shape S is very detailed, which can also increase the accuracy of face alignment. The operation method is shown in Formula (11).

$$S^{(t+1)} = S^{(t)} + r_t(I, S^{(t)}) \tag{11}$$

$$S = (x_1^T, x_2^T, \Lambda\ , x_p^T) \tag{12}$$

The face recognition process needs to use the training samples to compare the errors with their corresponding GroundTruths, so as to continuously optimize the training parameters. Face recognition using PCA first needs to establish a mean matrix, the calculation method is shown in Formula (13).

$$mean = \frac{1}{N} \sum_{k=1}^{N} I_k \tag{13}$$

Since it is an operation between matrices, the calculation method is shown in Formula (14):

$$A = [(I_1 - mean), (I_2 - mean), \Lambda\ , (I_N - mean)] \tag{14}$$

The face recognition system obtains the covariance matrix by multiplying the calculated difference matrix with its corresponding difference transposed matrix:

$$C = AA^T \tag{15}$$

Among them,

$$Dim(A) = (m * n) \times N \tag{16}$$

$$Dim(A^T) = N \times (m * n) \tag{17}$$

$$Dim(C) = (m*n) \times (m*n) \tag{18}$$

2.4. Evaluation of Algorithm Effect

From the perspective of the number of background updates, with the video frame length as the independent variable and the number of background updates as the dependent variable, Figure 5A shows the comparison of the number of background updates in the same video between the traditional method and the method in this paper, and the difference in the number of updates. It can be seen that as the video length increased, the number of background updates of the traditional method increased linearly, while the growth slope of the number of background updates using the method in this paper was much lower than that of the traditional method. Figure 5B can be obtained by curve fitting the difference between the two.

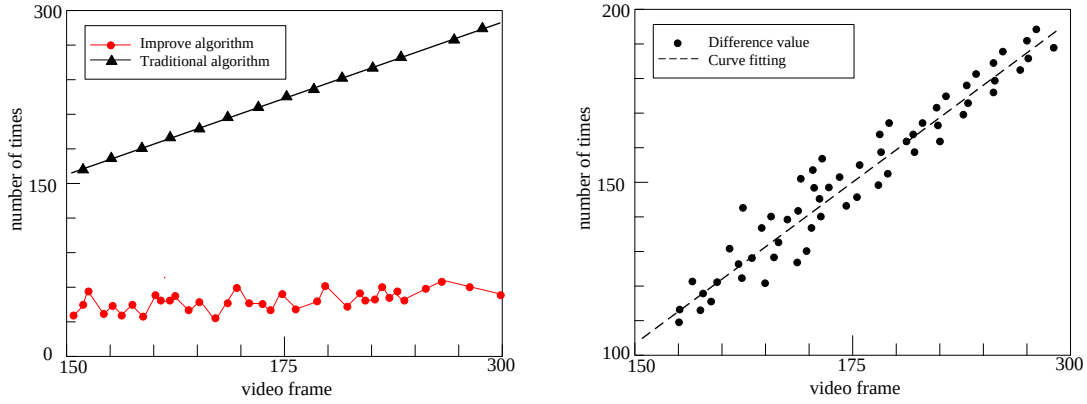


Figure 5. Comparison of different algorithms

From the perspective of algorithm running time, with the video frame length as the independent variable and the program running time as the dependent variable, the traditional background replacement method and the background replacement method of this paper were used for experimental comparison, and the results are shown in Figure 6. It can be seen from FIG. 6A that when the video length was short, the consumption time of the two methods was not much different. However, with the increase of video frames, the time advantage of the background update method adopted in this paper and the traditional method gradually increased. The time difference curves of the two methods showed an upward trend, as shown in Figure 6B.

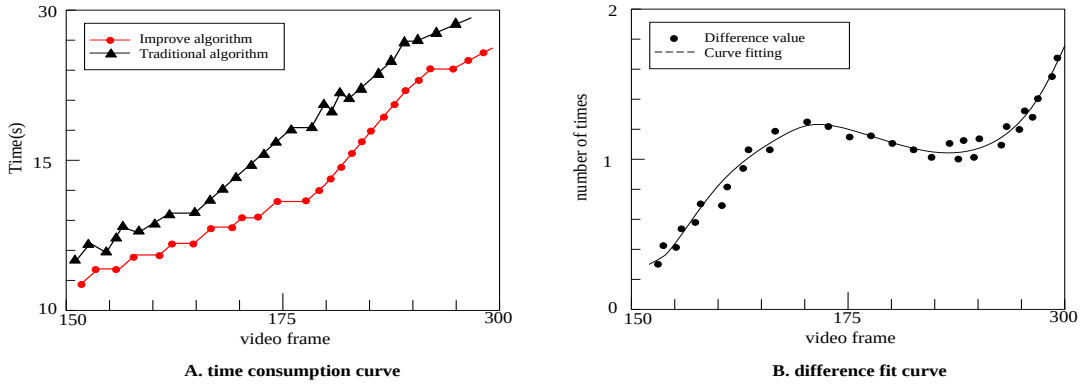


Figure 6. Computation time comparison

In order to further describe the effect of the traditional method and the method in this paper, the recognition rate of the algorithm was compared. In the experiment, 100 images of faces and hands were randomly obtained for recognition, and the results are shown in Table 1.

Table 1. Face and hand recognition accuracy

	Face	Hand
Traditional algorithm	73.8%	83.6%
Algorithm	90.1%	97.5%
Difference	16.3%	13.9%

As shown in Table 1, the recognition accuracy of the algorithm in this paper for face and hand was higher than that of the traditional algorithm.

3. National Vocal Music Teaching Methods

The teaching of national vocal music includes knowledge and technology. After decades of precipitation and development, the teaching of Chinese national vocal music has formed a scientific and systematic system [19]. In order to further study the current situation and characteristics of the teaching methods of ethnic vocal music, this chapter focused on the teaching methods of ethnic vocal music, and focused on analyzing the teaching of lyrics, rhythm, and rhythm in the teaching of ethnic vocal music.

3.1. Teaching of Lyrics in National Vocal Music

Chinese is very recognizable, and the pronunciation of each character can show the content and feelings it wants to express. The vividness of Chinese expression and the broadness of content provide specific and vivid lyrics for music singing and performance.

Chinese sounds can be divided into two categories: Ping and Zhe. The Ping-sound language is generally used to express things such as wide, empty, long, and sum; the Zhe-sound language is generally used to express narrow, solid, short, and urgent things. Combined with the four tones of Ping, Shang, Qu, and Enter or Yin, Yang, Shang, and Qu, the sounds of Yin Ping and Shang Ping are all Ping, and the rest are Zhe. Clear Chinese sound classification can distinguish the nature of lyrics in vocal works, so as to grasp the emotion of lyrics, and then determine the expression and action of singing performance [20].

Human emotions can be vividly and prominently displayed through language, and different emotions have different voice performances. Therefore, the emotion conveyed by the song can be expressed through the sound of the language. For example, the word "yo" in the first sentence of the song "Flower Girl", "Sell flowers come, sell flowers come" is a "go sound", in order to show the miserable and sad feeling of the flower girl, when singing, it is necessary to lengthen the pitch. The "mother" in the sentence "heal a sick good mother" is the word for "Ping-sound". When singing, it should be more gentle and full of love, showing the love and inner warmth of the flower girl when she thinks of her mother. The following is a detailed analysis of the expression of human emotions in the lyrics:

The first is when the characters are in a cheerful and happy mood. This requires that when uttering a voice, the mouth should be opened wide, and the voice will be more long and joyful. The second is the joy of the characters. Each character must be in a correct tone, the tone must be raised, the tone must be stably dropped, and sung completely, so that the joy of the whole piece can be vividly displayed. The third is the character's anger. The voice must be particularly loud and high-pitched when it comes out, and it must be resolute to show that it is necessary to use ones own

strength. Every word is sonorous and powerful, and it falls to the ground with a sound, which is the anger shouted from the bottom of heart. The fourth is the emotion of love. When expressing feelings of love, both actions and words must be gentle and long. The singer's mouth does not need to be opened very wide, and the voice comes out softly and softly, expressing in a long and delicate tone [21].

3.2. Teaching of Rhythm in National Vocal Music

Rhythm refers to all time-related factors in music, such as bars, beats, accent positions, cyclic term markers, etc. When a person can fully grasp and use these terminology factors to make the musical composition full of beauty and artistry, his musical rhythm must be good. Each musical work has two rhythm modes: one is the musical rhythm created by the composer; the other is the rhythm form presented in the second creation process of the singer during the singing process. This kind of rhythm is often a free rhythm, which is an important means of the secondary creation process, and it can also better show the singer's true emotions expressed by the singer from himself [22].

The first is the rhythmic style of the composer's work, which is often closely related to the content of the work itself. The small bridges and flowing waters in the land of fish and rice in the south of the Yangtze River, the smoky deserts in the Loess Plateau in the northwest, the strange peaks and scenery in the Yunnan-Guizhou Plateau and the vast plains in the northeast have different regional characteristics and language habits, all of which will affect the rhythm and form of the composer's musical works. The second is the free rhythm used by the singer for secondary creation when singing a song. The free rhythm is more flexible, the forms are diverse, and the creation is convenient.

3.3. Teaching of Rhythm in Ethnic Vocal Music

A beat is a combination of strong and weak beats. On the score, mark the rhythm with time signatures and bar lines, and the beat can be seen intuitively. Accurately grasp the rhythm, can quickly become familiar with the music, and feel the emotions conveyed by the music. The strong and weak beats in music mainly have distribution rules, as shown in Table 2:

Table 2. Beat distribution law

Beat	Type	Strong or weak
2 beats	2/2,2/4,2/8	Strong - weak
3 beats	3/2,3/4,3/8,3/16	Strong - weak - weak
4 beats	4/2,4/4,4/8	Strong - weak - secondary strong - weak
6 beats	6/4,6/8	Strong - weak - weak - second strong - weak - weak

The distribution law of the phrases determines the intensity distribution of the beats. For example, in the first paragraph of "Defending the Yellow River", the rhythm is tight, the rhythm is strong, and the tone is loud. Several factors together vividly depict the momentum of the Yellow River. The use of beats made people feel the turbulent waves of the Yellow River, the intensity and passion of resisting Japan, and the bravery and fearlessness of the guerrillas [23].

4. Design of Personalized Teaching System for Learning Resources

4.1. Demand

The teaching system developed based on the C/S architecture requires the user to download the client program before using it, and after the system is upgraded, the user must re-download and install the client before continuing to use it, so the later maintenance and upgrade process is cumbersome. In the system based on the B/S structure, users do not need to install special applications, but only need to use a browser to access. The security access requirements of multiple users can be achieved by setting access rights, and the maintenance and upgrading of the system are more convenient. At the same time, the teaching assistance system based on the B/S architecture has good cross-platform performance. Users can access the system through computers or mobile smart terminals, and experience augmented reality through terminal camera equipment.

4.2. Overall System Architecture Design

The interactive augmented reality education and teaching assistant system based on B/S researched and designed in this paper, the overall architecture consists of server, client (browser) and database. The server side is built on the Web server (WebServer), using Apache's lightweight application server Tomcat, and the server port is 8080. The server program is rapidly developed based on the Java language SpringBoot framework, and the SpringBoot framework is developed using the Controller-Service-Mapper three-tier architecture model. Responsible for processing requests sent by browsers, accessing MySQL databases, completing data management operations, and transmitting relevant results to client pages. The client interface is implemented by the Vue front-end framework. The front-end display part adopts HTML, CSS, JavaScript and iView component library, which can design a user-friendly page layout. Communication between client and server is based on Axios, an HTTP library that can be used in browsers and nodes. In js, it can choose post or get to communicate and transmit data information. The database design is based on Navicat's visual database management platform to complete the creation and data management of 3D model file information tables, user registration information tables and other data tables. The system completed the development and testing, and built the overall architecture through the IntelliJIDEA platform and the WebStorm platform to realize the augmented reality education teaching assistant system. The overall architecture of the system is shown in Figure 7.

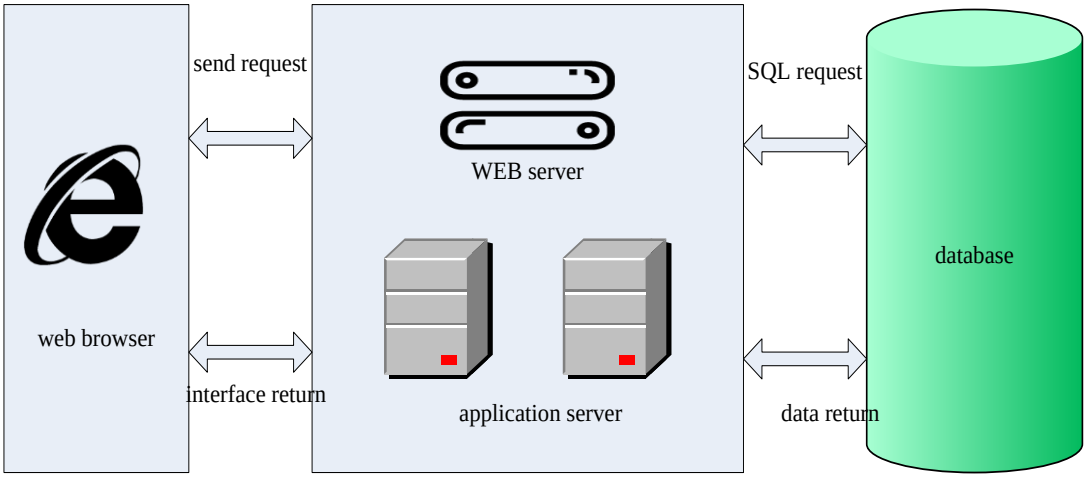


Figure 7. System architecture diagram

4.3. System Function Module Design

The system divides registered users into teacher user roles, student user roles and administrator roles. Through the specific analysis of system function permissions, all users of the system have basic functions such as registration and login, password modification and so on. Among them, teacher users have the authority to manage and maintain 3D model files, while student users have the authority to view 3D models under their teachers. In addition, both the teacher user and the student user have the right to apply for and unbind each other; the special authority of the administrator is the management of all registered user information, the management of 3D model data, and the system security management. The utilization of the 3D model resources of vocal music resources is shown in Figure 8.

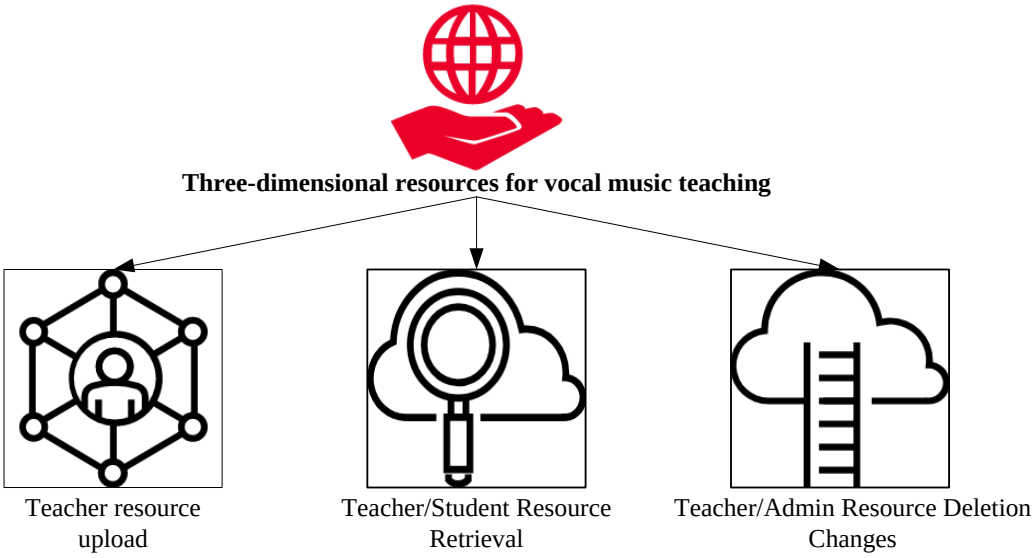


Figure 8. Utilization of vocal music teaching resources

Teacher users have the permission to upload 3D models, and the files to be uploaded include 3D model files and model texture pictures. The system judges the format of the locally uploaded file on the front-end page. If the format verification passes, it requests the file pre-storage path to the server. After the user submits and saves it, the server stores the file according to the pre-storage path, and the 3D model is uploaded successfully.

The 3D model file query is designed to facilitate users to quickly find and locate model files in the model file list. The query function module is designed in the system, and two query conditions are provided to narrow the search range. The user can enter the corresponding template name keyword and creation date range in the interface. The system searches the database according to the input keywords and date content. If the information of the template file is queried, it will display all the information on the web page.

In the practical application of the system, teachers can manage the uploaded 3D model files accordingly, including data modification and deletion. The modification of the 3D model file can be done on the edit page. The modification logic is similar to the upload logic, the data of the model file can be updated after saving; the deletion of the 3D model file can be realized on the model list page, and the user can select the data to be deleted and logically delete it.

4.4. Database and Server Design

In the teaching aid system designed in this paper, the design of the database is also an important part, and the database provides the necessary data support for the whole system. The MySQL database selected in this system is a lightweight and convenient relational database.

The development of the server side uses the mainstream Java language, and the overall architecture design is based on SprmgBoot, which simplifies the construction process of Spng applications and can quickly create Spring applications. The server provides an API interface to the outside world. External access can call the corresponding AH interface to transmit data to this interface in a certain format. The data processing on the server side adopts the three-layer design pattern of Controller, Service and Mapper. The Controller layer provides an interface for the outside world, and passes the incoming data to the Service processing layer; the Service layer mainly provides specific processing methods, performs corresponding operation processing on the data passed in by the Controller layer, and returns the result to the Controller layer. If need to operate the database, it can find a Mapper layer to execute the SQL statement; the Mapper layer supports addition, deletion, modification, and query operations on a single database table, and returns the processing results of the database table to the previous level. The relationship diagram of Controller, Service and Mapper is shown in Figure 9.

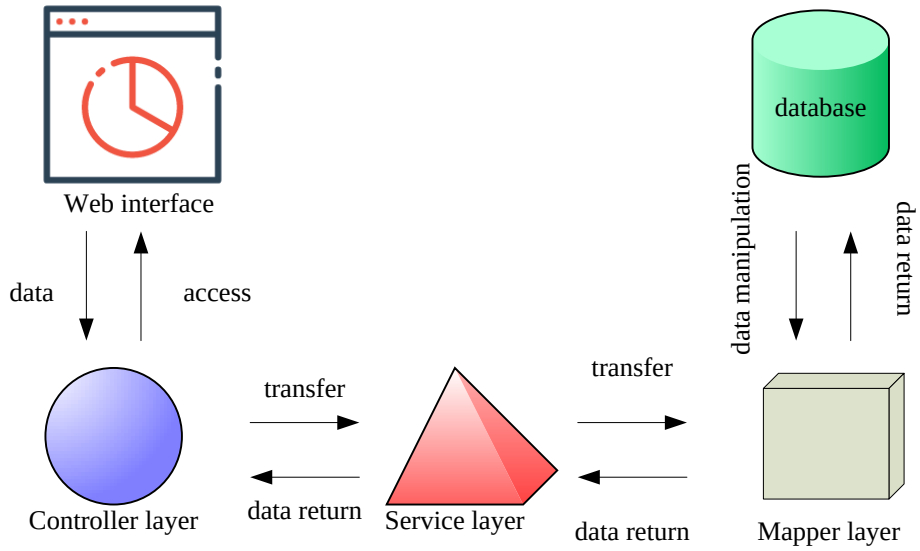


Figure 9. Multi-layer relationship diagram

4.5. System Function Test

After the system development is completed, the module function test of the system needs to be carried out to ensure the stable operation of the system in the actual application process. Both the client and server used in the test are 64-bit Windows PCs and CMOS cameras. In this paper, the system was tested by the black box test method, data operations were performed on the corresponding modules, and the stability of the entire system was obtained by recording the response results of each module. The test results of each functional module of the system are shown in Table 3 and Table 4.

Table 3. System functional module tests

Serial number	System module	Testing frequency	Success rate
1	Register module	100	100%
2	Login module	100	100%
3	Change password	100	100%
4	Teacher-student binding	100	99%
5	Model management	100	100%

Table 4. Computer vision functional tests

Serial number	System module	Testing frequency	Success rate
1	Hand moves up	100	97%
2	Hand down	100	99%
3	Move the hand to the left	100	96%
4	Move the hand to the right	100	96%
5	Face recognition	100	90%

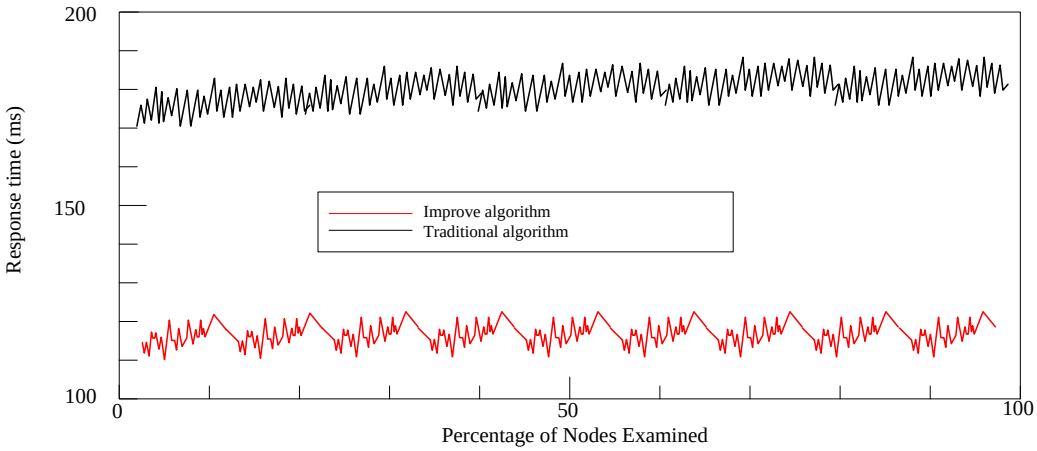
It can be seen from Table 3 that the system test results showed that the system has high stability in data operation and management performance, which meets the requirements of system stability. For the test of the augmented reality demonstration module, this paper conducted multi-model tests, and recorded the success rate of virtual fusion and human-computer interaction performance.

In this paper, according to the system's average response time to gesture actions, the system real-time test results were obtained. The test results are shown in Figure 10.

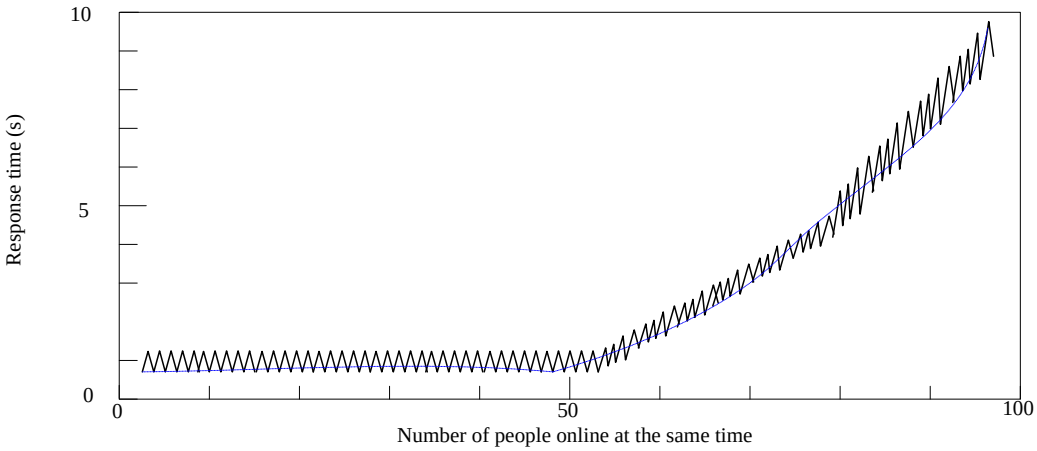
It can be seen from Figure 10A that the average response time of the system in the traditional method was about 178ms, and the average response time in this method was about 117ms, and the average difference between the two was about 60ms. Therefore, the adaptive background update method based on the increasing slope of the number of target pixels adopted in this paper can improve the real-time performance of the system to a certain extent.

It can be seen from Figure 10B that as the number of people online in the system increased, the response time of the system increased exponentially, but when there were less than 50 people, the response time remained at a good level. Therefore, the system of this paper can fully play a better function in ordinary classroom teaching.

This paper firstly described the system development environment, and introduced the realization results of each main module in the system. Secondly, the augmented reality demonstration module in the system was introduced in detail, including the training and use of artificial signs, and the introduction of human-computer interaction. Finally, each module of the system was tested respectively, and the test results showed that the system has high stability. In the augmented reality demonstration module, the virtual reality fusion and human-computer interaction performance meet the needs of the system.



A. The response time of the system applying different algorithms



B. The relationship between the number of people online at the same time and the response time

Figure 10. System performance test results

5. Conclusion

Computer vision technology can create a three-dimensional learning environment for students, which can bring a good sense of immersion to students and attract students to devote themselves to learning. Computer vision technology can help students understand the knowledge model and help students build spatial imagination. However, the development of teaching systems based on the Internet and computer vision technology is relatively slow, and the operation of the system is mostly in stand-alone operation mode, not integrated with the Internet. At the same time, the system does not have the human-computer interaction function based on natural gestures, which limits the application scope of the augmented reality teaching system to a certain extent. Therefore, the research and innovation of the teaching system based on the Internet and computer vision technology in this paper has positive significance in the field of education. However, there are still some problems in the setting of the function modules of the system. Due to the limited time, the content setting of each module in this system is not sufficient and needs to be further improved. In addition, for the security of the system, there is no security setting. In the future work, the security of the system will be further improved.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Lu K. National Music Promotion and Inheritance Strategies Based on the Perspective of Intangible Cultural Heritage. Arts Studies and Criticism, 2022, 2(4):46-52.

[2] Zhou Y. Discussion on the Theoretical Guidance and Practical Exploration of the Integrated Development of Aesthetic Education and Moral Education in Colleges and Universities. Journal of Educational Theory and Management, 2020, 4(2):7-13.

[3] Lei F, Wang X. Harmonic template neurons in primate auditory cortex underlying complex sound processing. Proc Natl Acad Sci U S A, 2017, 114(5):E840-E853.

[4] Liu X. The narrative in vocal music: the genre aspect. Aspects of Historical Musicology, 2019, 18(18):230-242.

[5] Kang H. Analysis on selection of vocal music works and effective teaching methods in vocal music teaching in colleges and universities. Revista de la Facultad de Ingenieria, 2017, 32(15):280-283.

[6] Li Y, Zhu J. Research on the personalized teaching system of college English. Agro Food Industry Hi Tech, 2017, 28(1):2530-2534.

[7] Zang R, Wang L. Personalized Teaching Model of College English Based on Big Data. Journal of Physics Conference Series, 2021, 1852(2):022013-022025.

[8] Marinagi C, Skourlas C. Blended Learning in Personalized Assistive Learning Environments. International Journal of Mobile & Blended Learning, 2017, 5(2):39-59.

[9] Zhao, JC, Chen, Guo, JX. A recommendation algorithm based on Collaborative Filtering Technology in distance learning. Electronic and Automation Control Conference (IAEAC), 2017, 2017(-):2320-2323.

[10] Tang Y, Liang J, Hare R, Wang F Y. A Personalized Learning System for Parallel Intelligent Education. IEEE Transactions on Computational Social Systems, 2020, 7(2):352-361.

[11] Sunkara V M, Kurra R R. An Analysis of Learner Satisfaction and Needs on E-Learning Systems. International journal of computational intelligence research, 2017, 13(3):433-444.

[12] Zhang Z, Min H. Analysis on the Construction of Personalized Physical Education Teaching System Based on a Cloud Computing Platform. Wireless Communications and Mobile Computing, 2020, 2020(3):1-8.

[13] Xu Y. Computer-Aided Design of Personalized Recommendation in Teaching System. Computer-Aided Design and Applications, 2019, 17(S1):44-56.

[14] Rusdiana A. Personalized System of Instruction (PSI) Models: Using Digital Teaching Materials on Learning. Creativity and Innovation Management, 2020, 9(5):314-324.

[15] Arthur Damio Médici, Carmo J D S. Personalized System of Instruction (PSI) and Lesson Study (LS): a comparative analysis of two successful teaching methods. Science and Knowledge in Focus, 2018, 1(1):05-11.

[16] Wang J. Research on College English Teaching System Based on Computer Big Data. Journal of Physics: Conference Series, 2021, 1865(4):042141-042151.

[17] Gai F. Construction of ESP Teaching System. Journal of Language Teaching and Research, 2017, 8(6):12041209-12041235.

[18] Gan B. Design and Application Research of VR/AR Teaching Experience System. Journal of Physics: Conference Series, 2019, 1187(11):052079-052088.

[19] A L V, B G D P, B G S. A methodology for voice classification based on the personalized fundamental frequency estimation. Biomedical Signal Processing and Control, 2018, 42(38):134-144.

[20] Yang Z. The challenges of personalized learning and their solutions. Chinese Journal, 2019, 64(5-6):493-498.

[21] Guo Q. Physical education multimedia teaching model based on the fusion of dynamic bee colony algorithm. Boletin Tecnico/Technical Bulletin, 2017, 55(12):136-144.

[22] Li T. Construction and Implementation of Network Teaching Platform for Design Art Education Based on Cloud Technology. Educational Sciences: Theory and Practice, 2018, 18(5):13-19.

[23] Tafazoli D, Parra E G, Abril C. Intelligent Language Tutoring System: Integrating Intelligent Computer-Assisted Language Learning Into Language Education. International journal of information and communication technology education: an official publication of the Information Resources Management Association, 2019, 15(3):60-74.