Scholar Publishing Group
International Journal of Educational Curriculum Management and Research
https://doi.org/10.38007/IJECMR.2023.040309
ISSN 2790-0967 Vol. 4, Issue 3: 83-97

Multimodal Music Teaching Mode Based on Human-computer Interaction Technology

Yanfang Ou*

Philippine Christian University, Manila, Philippines

982431208@qq.com

**corresponding author*

Keywords: Multimodal Teaching Mode, Human-Computer Interaction Technology, Music Teaching, Kinect Somatosensory Technology

Abstract: Music teaching is an activity completed by both "teaching" and "learning". Therefore, whether the two sides of the teaching can achieve close communication and cooperation has become a very important factor in determining the quality and efficiency of music course teaching. The multi-modal teaching mode integrates multiple modes, fully mobilizes students' various senses, attracts students' attention, makes students more relaxed and easy to grasp and understand the knowledge they have learned, and achieves better teaching effects. In the multimodal music teaching class, teachers make full use of ppt courseware to play audio and video, display pictures and tables, use artistic words, font colors and underlines, and use rich body language to enable students to learn language and music knowledge in vivid teaching activities. Human-Computer Interaction is a key technology in the field of virtual reality. Its main purpose is to improve the interaction between the user and the computer from the user's point of view, thereby increasing the immersion and authenticity of the system. In human-computer interaction, human behavior recognition is a very critical technology, which helps human beings to achieve a natural and harmonious state of human-computer interaction. This paper studied the multimodal music teaching model based on Kinect somatosensory technology. It was found that most of the indicators of the experimental class using the multimodal music teaching model are better than the control class, and the average scores of the experimental class are all more than 85.5. It showed that the use of multimodal teaching mode to teach can enhance students' classroom participation, help students understand and memorize knowledge, and stimulate students' interest in learning.

1. Introduction

Human-computer interaction is gradually becoming more human-centered and improving the user experience. Whether it is interactive input or display output, interaction design pays more

attention to the user's feelings, and the user-oriented user experience can reduce the user's cognitive burden and learning cost in the interactive process, creating a natural and full-hearted interactive environment. Especially in recent years, with the rapid development of technologies such as virtual reality, augmented reality, and somatosensory interaction, perception technology has become the focus of competition among major technology companies in recent years. The multimodal teaching mode is developed on the basis of the multimodal discourse analysis theory, and it is mainly applied to English teaching at first. Music education can improve and cultivate people's quality, and it has unique functions and functions. Therefore, music education in the new era will face many new reforms, otherwise China's music education will always remain in the traditional form of music education. The music teaching model closely links theory and practice. It not only concretizes teaching theory, but also systematically summarizes teaching practice experience. It has important practical significance to study how to fully reflect the superiority of the music teaching mode.

With the rapid development of science and technology, the society has entered the information age, and the multimodal music teaching model that fully mobilizes students' multi-sensory participation in learning has gradually become the focus of research in the field of education. Fortuna S reflected on the question of whether a greater number of multi-sensory and incarnational approaches could re-define some hypotheses about music teaching. Recently research on embodiment and multimodal perception in instrumental music learning and teaching could suggest new directions for music education. The idea of considering whether a combination of activities such as body movement, listening, visualization of mimics and singing could be more effective for the processes of musical intelligibility than a decorporeal and sporadic approach has great relevance in practice [1]. The worldwide outbreak of COVID-19 has had a particularly widespread affect on educational systems. In the epidemic situation, the teaching of the English as a foreign language for adults should be integrated with the characteristics of education in English and the needs of adult learners. Huang R was based on multimodal discourse analysis theory. The multimodal teaching model for adult English learners guided teaching using a multimodal learning platform, carried out on-site teaching, and organized trimodal teaching activities from three levels of teaching preparation, teaching organizing, and teaching evaluation, in order to stimulate adult learners' enthusiasm for learning English, develop their autonomous learning ability, and contribute meaningfully to adult learners' constructing and internalizing themselves knowledge [2]. Education informatization has become a new trend in the "Internet +" era. English education in China still follows the traditional teaching system, which pays more attention to the imparting of knowledge rather than practice and training. The establishment of MOOC, the multimedia teaching of micro-courses, and the network public platform have realized the combination of online and offline teaching modes. However, in the process of using modern information technology, most colleges and universities have not formed a systematic model. Therefore, it is difficult to reach the level of modern teaching in English teaching. Based on the information technology of "Internet +", Lin W proposed an innovative method of multi-modal English teaching system integrating text mode, image mode and regression mode, aiming to design a more reliable theoretical paradigm for college English teaching. A reference for colleges and universities was provided to expand the network teaching mode [3]. Shustrova E V shared information on the ways in which different kinds of polymodal or so called kriolized texts can be a very useful means of studying theoretical topics. The main focus is on what traditionally belongs to the future area of specialization in modern foreign language teaching. Typically, these theoretical topics have a wealth of linguistic and cultural facts and phrases that shall be presented and exposed to students. They described how purely informative linguistic facts can be presented and are taught by means of graphics. The method of investigation was based on both conceptual metalogical theory and multimodal metaphorical methodology. Multimodal texts implied the use of both verbal and visual components. Thus, different advertising

techniques, comics (including political topics), comic strips, posters, collages and feature films can present an infinite number of sources of material. The basic results may be of interest to workers and scholars working with linguistic and cultural statistics [4].

Human-computer interaction technology has always been an important research topic and one of the research directions of artificial intelligence development. Among them, gesture recognition is very intuitive and natural, and it is the research focus of human-computer interaction. Yu F X built an intelligent gesture recognition framework based on the sensor Kinect, and improved the optimization algorithm. First, the final number of fingertips by analyzing the contour and the shape of the palm by determining the number of fingertips were determined, and the shape of the final gesture by comparing the shape of the palm before and after is determined, so as to realize the action that the operator wants the device to achieve. Finally, through example analysis, it can be seen that this method had a good recognition rate [5]. The teaching methods with the help of information technology have been rapidly growing under the concept of an information freeway. Based on the constructivist study theory, Liang W has proposed a new teaching method, including reforming teaching contents and reforming teaching methods. It is an inevitable trend of current education development to realize the informationization of education and meet the challenges to education in the information age. As the second generation web language, the method of VRML was introduced. Combining the feature of actual cases, the application of VRML in virtual learning circumstances was analyzed, and the specific implementing method was given. For the use of virtual reality modeling language (VRML) in online teaching, the basic theory of CAI and principles of online teaching were discussed. The web-based teaching mode based on VRML was divided into two application modes, and VRML-based web teaching mode was proposed [6]. In recent years, vision-based gesture recognition (VGR) has become a research hotspot in the field of human-computer interaction (HCI). GAO reviewed state-of-the-art vision-based approaches from different stages of the gesture recognition process, namely, (1) image acquisition and preprocessing, (2) gesture segmentation, (3) gesture tracking, (4) feature extraction, and (5) gesture classification. The state-of-the-art vision-based gesture recognition methods were reviewed. He also analyzed the advantages and disadvantages of various methods in detail. Finally, challenges and future research directions for vision-based gesture recognition in tactile rendering were discussed [7]. These methods provide some references for our research, but due to the short time and small sample size of the relevant research, this research has not been recognized by the public.

This paper put forward the necessity of establishing a multi-modal music teaching model based on artificial interaction technology, and conducted experimental research. Using the palm position and attitude data obtained by Leap Motion for filtering, except for the first 50 data points, the Kalman filtering algorithm effectively smoothed the gesture data, which showed that it is necessary to set the start and end points of gesture recognition. In order to understand the current situation of music classroom teaching mode and the necessity of improvement, a field survey was conducted, and it was found that 71.7% of students and 71.4% of teachers believed that the level of interaction and cooperation between teachers and students in the classroom was average or not good, indicating that traditional music teaching needs to be improved. Two classes were selected for comparative experiments. Most of the indicators of the experimental class were better than those of the control class, and the average scores of the experimental class were all more than 85.5. In addition, the frequency of using multimodal resources in classroom teaching was relatively high, and the frequency of using text and pictures was as high as 90%.

2. Method of Multimodal Music Teaching Mode Based on Human-computer Interaction Technology

Integrating virtual reality and human-computer interaction technology can not only allow people to roam in the virtual world, but also obtain a near-real interactive experience, and experience more interaction in the virtual world [8]. The introduction of virtual reality technology into human-computer interaction technology will make the ways of human-computer interaction more diversified. People can use their limbs to operate, and computers can also achieve 3D immersion through virtual reality devices such as head-mounted displays and ring screens [9]. They are embodied in:

(1) The interactive environment is more realistic, and the operator is as indistinguishable as in the real world when interacting with the computer;

(2) The interactive devices are more abundant. With the emergence of helmets, data gloves, position trackers, data clothes, and 3D glasses, operators can make full use of the hands, head, and eyes of the human body to interact, which greatly improves the the immersion of the interaction [10].

(3) The interactive interface is more friendly. Human-computer interaction based on virtual reality technology is closer to the real multi-channel, human-centered human-computer interaction, which is the development direction of future human-computer interaction technology [11]. The composition of the human-computer interaction system based on virtual reality is shown in Figure 1.

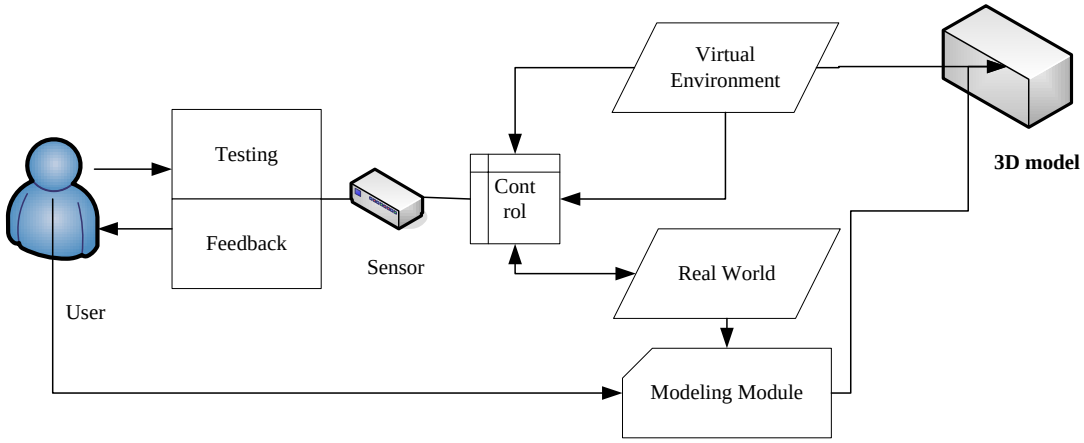

Figure 1. Composition of the human-computer interaction system based on virtual reality

Using a variety of sensors, it is very convenient to use the activities of the body to interact with the surrounding instruments or the surrounding environment. It allows users to interact with surrounding objects without resorting to other complex operating devices [12]. With the rapid development of technology, the expectations for the human-machine interaction are becoming higher and higher, and more and more investigators are beginning to focus on new techniques of interaction such as manual gesture input, speech detection and sensual feedback [13].

Kinect is a new body-sensing device that enables relevant interactions through instant motion, facial recognition and voice recognition. This feature makes Kinect make a qualitative leap in human-computer interaction technology, and promotes the development of Kinect to a certain extent [14].

The Kinect device is different from the traditional input device. It allows users to control the terminal through their limbs, and conduct human-computer interaction in the most natural way, which can not only meet the natural needs of people, but also promote the development and

promotion of human perception technology [15]. Table 1 compared the performance of Kinect1.0, Kinect2.0 and Leap Motion.

Table 1. Performance comparison of two somatosensory sensors

Body Sensor	Kinect1.0	Kinect2.0	Leap Motion
Resolution	320*240	512*484	640*240
Monitoring maximum range	4m	4.5m	600mm
Accuracy	4mm	1mm	0.01mm
Number of recognized joints	20	25	27

Throughout this article, Kinect was employed to recognise and keep track of people targeted in a scene. When many targets are simultaneously present in the work area, it is required to be able to recognize designated persons and perform location tracking quickly and exactly, while non-target persons are not recognized [16]. Target identification and tracking mainly includes target persons identification and target persons positional tracking. Figure 2 shows the whole process of target identification and target tracking.

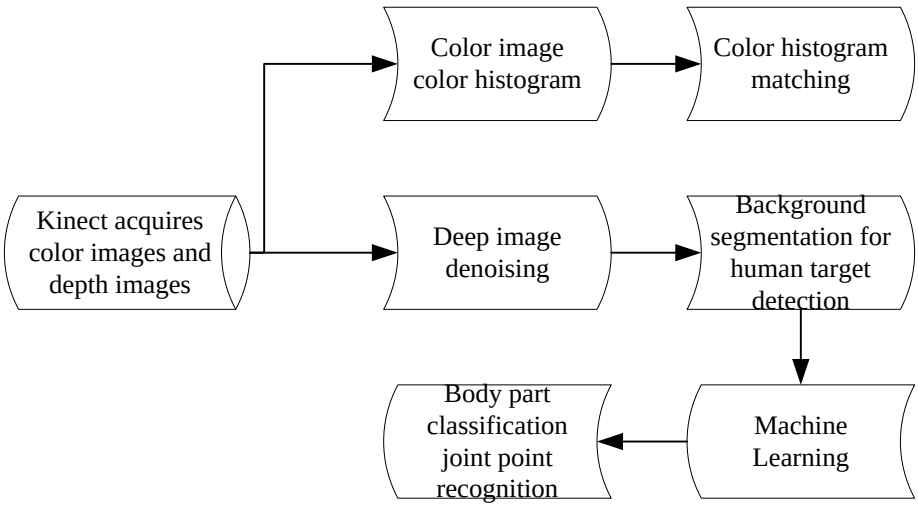

Figure 2. Object recognition and tracking process

In this paper, an imaging based person awareness method is employed to capture the color image of the scene by Kinect and then identify the target person in the scenario from the matching resulted of the color the histogram. The color of histogram is mainly used to describe the proportion of dissimilar colors in the whole image [17]. The common color of color histogram matching methods mainly include correlation method, card method and cross over method. The calculation formula of correlation is illustrated in Formula (1).

$$d_c = (H_1, H_2) = \frac{\sum_i H_1(i)H_2(i)}{\sqrt{\sum_i {H_1}^2(i){H_2}^2(i)}} \quad (1)$$

$$H_c' = H_c(i) - \frac{1}{N}(\sum_j H_c(j)) \quad (2)$$

Among them, N represents the number of equal divisions of H components, and H_1 and H_2 are histogram functions.

The formula for calculating chi-square is:

$$d_{ch} = (H_1, H_2) = \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)} \quad (3)$$

The formula for calculating the intersection is:

$$d_i = (H_1, H_2) = \min \sum_i H_1(i) H_2(i) \quad (4)$$

According to Formula (1), Formula (2), Formula (3) and Formula (4), the matching result comparison table shown in Table 2 was gained. The greater the degree, the greater the value of 1 for exact matched. The smaller the calculation result of the chi-square matching algorithm, the higher the matching degree of its histogram, and the value is 0 when it is completely matched.

Table 2. Histogram matching method reference table

Matching results	Complete mismatch	Semi-Match	Exact match
Correlation	-1	0.75	1
Cardinality	2	0.67	0
Intersection	0	0.5	1

In this paper, the method of color histogram matching is used to identify the target person. When performing target recognition, the color image of the target person needs to be collected in advance, and the color histogram of the target person is obtained through image processing and stored in the system as a template for target recognition [18]. The target recognition process is shown in Figure 3. The color image of the person in the scene is obtained through Kinect and its color histogram is obtained through image processing. The color histogram is then matched with the color histogram template stored in the system to identify the target person [19].

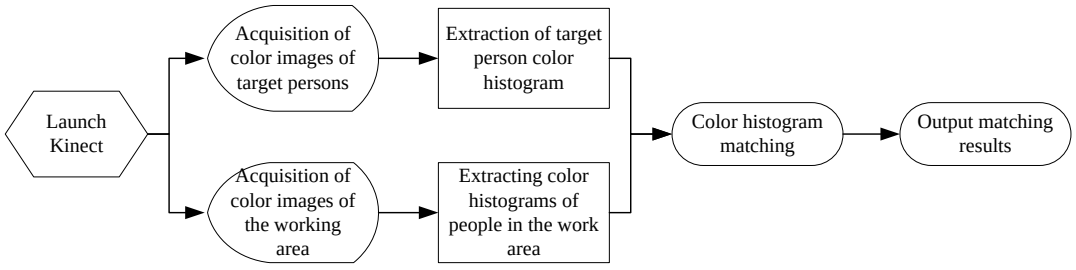

Figure 3. Object recognition flowchart

The action recognition results of the underlying HMM were used as the observed series of the high-level instructional path selection HMM to train the recognition of the multi-instructional protocol model [20]. The observing sequence of the base layer identifications can be indicated as follows.

$$s(t) = \{s^1(t), s^2(t), \ldots, s^N(t)\} \quad (5)$$

In the formula, $s^i(t)$ represents the digital number of the recognized action.

The forward variable and backward variable formulas in the discrete HMM model are modified to get the formula:

$$a_{t+1} = \left[\sum_{i=1}^{k} a_t(i) a_{ij} \prod_{f=1}^{N} p_j(s^f(t+1))\right] \quad (6)$$

$$b_{t+1} = \left[\sum_{i=1}^{k} b_t(i) a_{ij} \prod_{f=1}^{N} p_j(s^f(t+1))\right] \quad (7)$$

In the formula, $a_t(i)$ is the output probability of the HMM model in the forward algorithm observing the partial observation sequence in a certain state at time t; $b_t(i)$ is the output probability obtained by observing part of the observation sequence of the HMM model at time t in the backward algorithm; $s^f(t)$ is the variable of the observation sequence.

The revaluation formula for confusion matrix B is modified as:

$$\overline{p_j^f}(k) = \text{correct}(k^f|j)/\text{correct}(j) \quad (8)$$

Among them, $\text{correct}(k^f|j)$ represents the correct number of actions in the action recognition sequence under teaching content s_j; $\text{correct}(j)$ represents the expected value of the correct number of actions in this action recognition sequence. In teaching practice, in order to enrich the teaching form, teachers skillfully adopt and rationally apply elements based on teaching objectives, teaching material content, students' performance and other subjects [21]. The music teaching mode provides teachers with a teaching frame structure that concretizes abstract theories, so that teachers' teaching behavior can be standardized. It can also present complex teaching theories in a simplified form, making it easier for students to master knowledge. Then it shows in the form of concise and clear graphics or tables and so on. The characteristics of the music teaching mode are operability, and the general teaching mode also has its characteristics.

The music teaching mode must take into account the characteristics of music itself, and the teaching mode cannot be discussed without music. For other disciplines, music is more unique and has its own regularity. Music is inseparable from sound. It is expressed through sound. Music is artistic, and this art is not only to express people and things, but more importantly, to highlight people's emotions and thoughts. Aesthetic and musical education through art are among the most important means. Music education is to cultivate students' creative thinking and rich imagination through the way students appreciate music, create music, and express music, so that students can develop a positive and healthy attitude, and cultivate their sentiments. It can be concluded that the characteristics of music teaching mode are musicality.

In the music class, not only students are required to respect their teachers, but teachers should also care about and care for students. However, students respect teachers, teachers should also respect students. They should always encourage students to show their talents bravely in teaching, and provide more opportunities for students to participate. In this way, students can discover their own shortcomings and develop the habit of learning from each other. Teachers cooperate with students to complete teaching activities, cultivate students' team spirit, and promote harmonious coexistence between teachers and students and between students. Through a series of teaching activities, students are constantly motivated to learn, and students can take the initiative to understand, explore and discover behaviors in music. In teaching, the teaching form needs to be open, and the teachers' teaching concept also needs to be open. It is concluded that the characteristic of music teaching mode is openness.

In the process of music teaching, there will often be some teaching activities, and teaching activities occupy an important position. Without teaching activities, it is conceivable that students cannot understand abstract teaching theories, and students will always feel that music is a boring subject. Different styles of music, different performance skills, different singing emotions, etc., all need teaching activities to distinguish. Teaching objectives can only be achieved through teaching activities. It can be seen that the characteristics of music teaching mode are activity.

The structure of the music teaching mode is carried out under the guidance of certain teaching

ideas, and is formed through a lot of teaching reflections. It takes the essence of various teaching practice experiences, so it can be seen that the music teaching mode has advantages. Therefore, it can be seen that the music teaching mode has superiority, but the general teaching mode does not exist, and the failed and backward teaching mode will be eliminated. For example, the traditional teaching model focuses on the status of teachers and ignores students. There is only a give-and-take relationship between teachers and students. When students master the knowledge, they will not eat it, and over time, students will refuse to learn music. In the end, the teaching effect cannot be achieved. Therefore, the characteristics of music teaching mode also have advantages.

"Modal" is a term used in computer science to study human-computer interaction. It refers to the comprehensive use of multiple modalities, through text, images, videos, gestures, three-dimensional animation and other means of communication and symbolic carriers. The multimodal teaching mode mainly uses the network platform, modern media technology, etc., through pictures, audio and video, physical objects, facial expressions, body gestures and other means to fully mobilize the learners' multiple senses to participate in classroom teaching. The traditional multimodal teaching mode is shown in Figure 4. Human-computer interaction technology emerged with the invention of robots. In the rapid development of robotics, human-computer interaction technology has also made great progress as an important part. The rapid development of information systems has facilitated the interaction between humans and machines, thus triggering extensive research on human-computer interaction technologies. Virtual reality is a technology that uses computer-generated three-dimensional visual effects, which can realize the movement of the virtual world according to the user's actions. Virtual reality technology provided users with a brand-new, brand-new, interactive and immersive experience, which greatly enhanced the viewing experience of the scene. It is a major breakthrough in human-computer interaction technology. Figure 5 is a multimodal teaching mode assisted by a multimedia device.

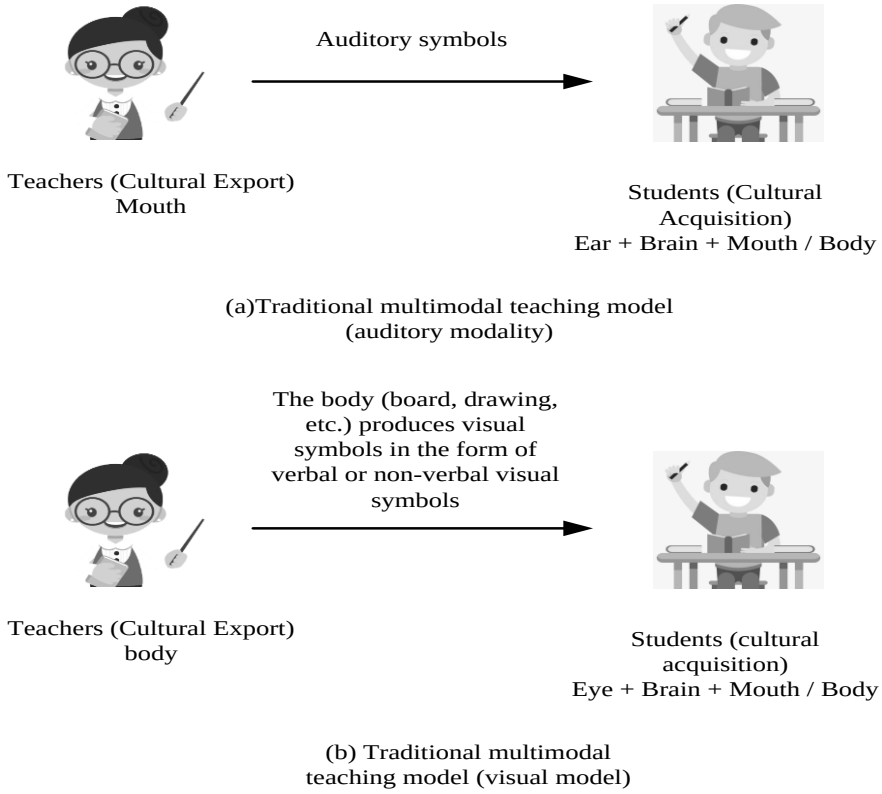

Figure 4. Traditional multimodal teaching model

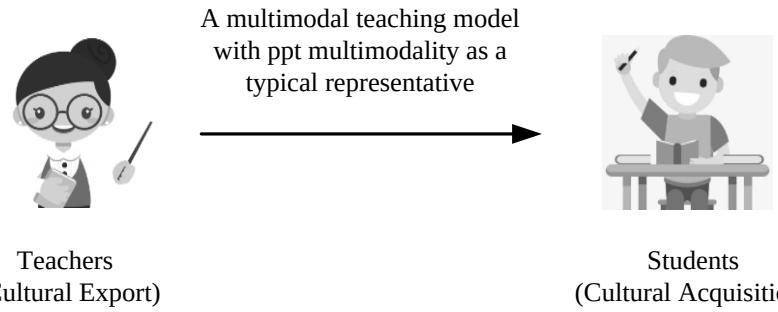

Figure 5. Multimodal teaching mode assisted by multimedia devices

Each modality type plays a role that cannot be ignored in teaching. Language is the most effective and direct way to transmit information. Therefore, the modalities are divided into two categories: one is language modalities (including spoken language and text, which are often used to express meaning and convey information), and the other is non-linguistic modalities (including auditory, visual, taste, smell, paralinguistic modalities, etc. Among them, visual modalities, auditory modalities, and paralinguistic modalities are often used in non-linguistic teaching). The modal types are shown in Table 3.

Table 3. Modal types

Serial number	Modal Classification	Types of Modality
1	Language Modality	Spoken language, written language, voice tone, accent
2	Body Modality	Gestures, body movements, head movements, expressions, eyes, orientation
3	Visual Modality	Images, animation, clothing, colors
4	Auditory Modality	Music, singing, rhythm
5	Environmental Modality	Classroom layout, ecology, physical distance

In the process of human action recognition, the foreground image can be obtained by the background clipping method first, and then the edge contour can be extracted by the Canny algorithm.

(1) Representation of contour map

Assuming that the number of horizontal pixels in the image is N, the number of vertical pixels is M, and the set of pixels in all images is:

$$T = \{(x, y) | 0 \leq x < N, 0 \leq y < M\} \tag{9}$$

Assuming that the gray value of any pixel point (x, y) is gxy, in the extracted contour map (as shown in Figure 2), the set of contour points is:

$$W = \{(x, y) | (x, y) \in T, gxy \neq 255\} \tag{10}$$

Then the set of non-contour points is:

$$V = \{(x, y) | (x, y) \in T, gxy = 255\} \tag{11}$$

(2) Representation of candidate regions

Now a pixel area of 3×3 is used to represent the candidate area, and this area is marked. The set of pixel points in the candidate area is:

$$r = \{(x,y)|(x,y) \in \cup_{i=0}^{8}\{r_i\}\} \tag{12}$$

Assuming that the coordinate of R_0 is (x_0, y_0), the coordinate position relationship between R_0 and $R_1, R_2, \ldots, R_8$ is as follows:

$$R_1: \begin{cases} x_1 = x_0 - 1 \\ y_1 = y_0 \end{cases}, \ldots\ldots, R_8: \begin{cases} x_8 = x_0 - 1 \\ y_8 = y_0 - 1 \end{cases} \tag{13}$$

(3) Representation of vectors

The 9 pixels in the candidate area can be expressed as 8 packets with R_0 as the starting point and $R_1, R_2, \ldots, R_8$ as the end point respectively. They are:

$$\underrightarrow{R_{10}} = R_1 - R_0, \ldots, \underrightarrow{R_{80}} = R_8 - R_0 \tag{14}$$

The vector set is:

$$Q = \left\{q|q = \underrightarrow{R_{i0}}, 1 \le i \le 8\right\} \tag{15}$$

In order to facilitate the calculation, each vector is recorded in the form of a column vector.

(4) Representation of centroid

The formula for calculating the centroid is as follows:

$$x_k = \frac{1}{N_b}\sum_{i=1}^{N_b} x_i \tag{16}$$

$$y_k = \frac{1}{N_b}\sum_{i=1}^{N_b} y_i \tag{17}$$

3. Experimental Design of Multimodal Music Teaching Mode

Music teaching involves playing a lot of musical instruments. In this paper, target recognition tracking and gesture recognition are mainly carried out through somatosensory devices. After confirming the operator's identity through Kinect, the gesture information is recognized by Leap Motion and used for robot control to realize natural interaction between human and robot. This control method combined the high efficiency and high precision of the robot with the natural and flexible characteristics of the human being.

Combined with the actual situation of the laboratory, this paper used ABB IRB1410 robot for experimental verification. The overall structure of the human-computer interaction system is shown in Figure 6, which mainly includes the somatosensory sensor, the PC and the IRB1410 robot. Communication between the PC and the robot was carried out through the C/S structure. The system functions as follows:

(1) The system can accurately identify the operator's identity and obtain its location information;

(2) Gesture data in real time is obtained, and the robot to operate through gestures is controlled directly;

(3) The client can feedback the pose of the robot in real time.

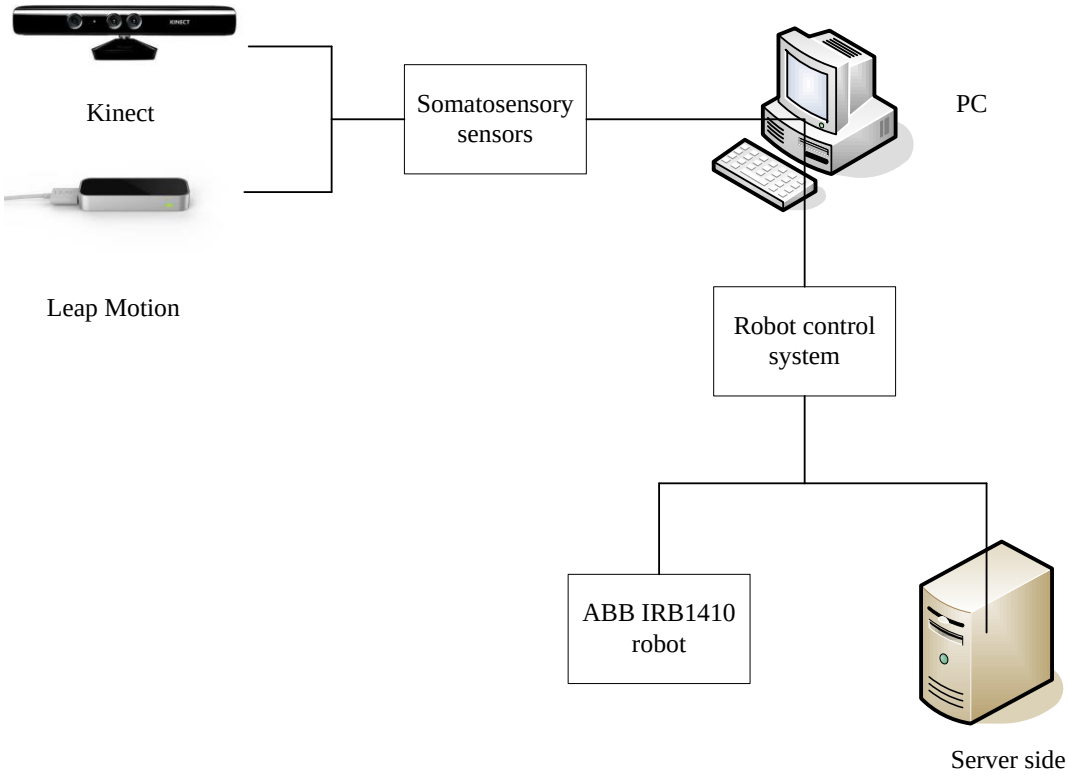

Figure 6. Overall structure of human-computer interaction system

When collecting gesture information, the operator places the hand in the recognition area of Leap Motion, and the device can actively obtain the hand information and obtain data information through the Leap SDK. Leap Motion acquires gesture information for the recognition area at 50 frames per second.

4. Experimental Data of Multimodal Music Teaching Mode

In this paper, the palm position and attitude data obtained by Leap Motion are filtered. The Kalman filtering algorithm can effectively remove jitter and mutation data, so that the acquired gesture data is more stable, avoiding the wrong operation of the robot due to the mutation of the gesture data, and ensuring the safety of the operator. Figure 7 is a comparison diagram of the position and attitude data before and after filtering. The black curve in the figure represents the unfiltered data sequence, and the red curve represents the filtered image sequence.

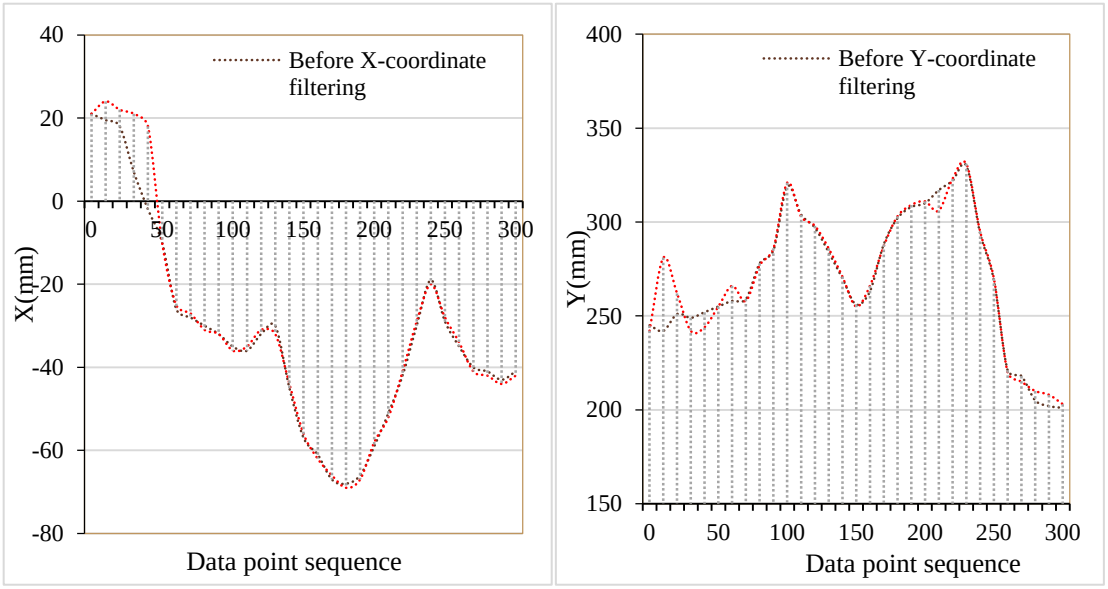

Figure 7. Comparison of filtering processing

It can be seen from Figure 7 that the Kalman filter algorithm effectively smoothes the gesture data. However, due to the large mutation of the data in the early stage of data collection, the collected position and attitude data were obtained directly without judging the radius of the palm ball, resulting in an unsatisfactory smoothing effect for the first 50 data points, which further illustrated the necessity of setting the start and end points of gesture recognition.

For a long time, under the shackles of traditional concepts, Chinese music teaching still follows the teaching method of "single indoctrination by teachers and passive acceptance by students". In order to further understand the current situation of music teaching, field research was conducted in the second grade (1) and (2) classes of a junior high school, and 120 questionnaires and questionnaires from 7 music teachers were distributed in the classroom. The teachers actively assisted, cooperated, collected all the questionnaires in the classroom, and verified that the obtained questionnaires were all valid. The following results were obtained, as shown in Figure 8.

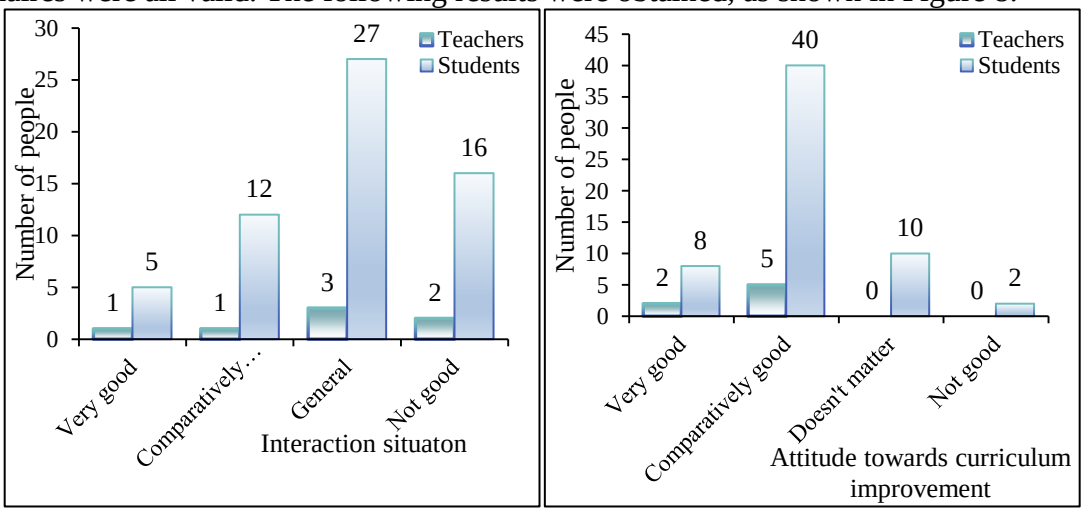

Figure 8. Statistical results on the interaction between teachers and students in teaching and the reform of the current music teaching methods

As can be seen from Figure 8, 27 students and 3 teachers think that the level of interaction and

collaboration between teachers and students in the classroom is average, and 16 students and 2 teachers feel bad, accounting for 71.7% and 71.4%, respectively. This shows that traditional classroom teaching lacks teacher-student interaction and student-student interaction, ignores students' participation experience, and is not conducive to mobilizing students' interest in the knowledge they have learned and the fun of participating in classroom teaching, thus affecting the teaching effect. When asked whether they want to reform the current teaching method of music class based on indoctrination, 80% of the students (48 people) answered yes, and only 2 students thought it was not good. This result is in line with expectations. This shows that students are generally bored with the traditional music teaching mode. According to the current teaching status, the characteristics of music and the specific conditions of students, it is necessary to actively explore a scientific and efficient teaching method.

In order to have a more comprehensive understanding of the current situation of music teaching in junior high schools and the feelings, expectations and requirements of music teachers in front-line junior high schools, (1) class is a control class, and the traditional teaching mode is still adopted; (2) class is an experimental class, which adopts a multi-modal music teaching model for teaching, and conducts an investigation after one month of teaching. The survey results are shown in Figure 9.

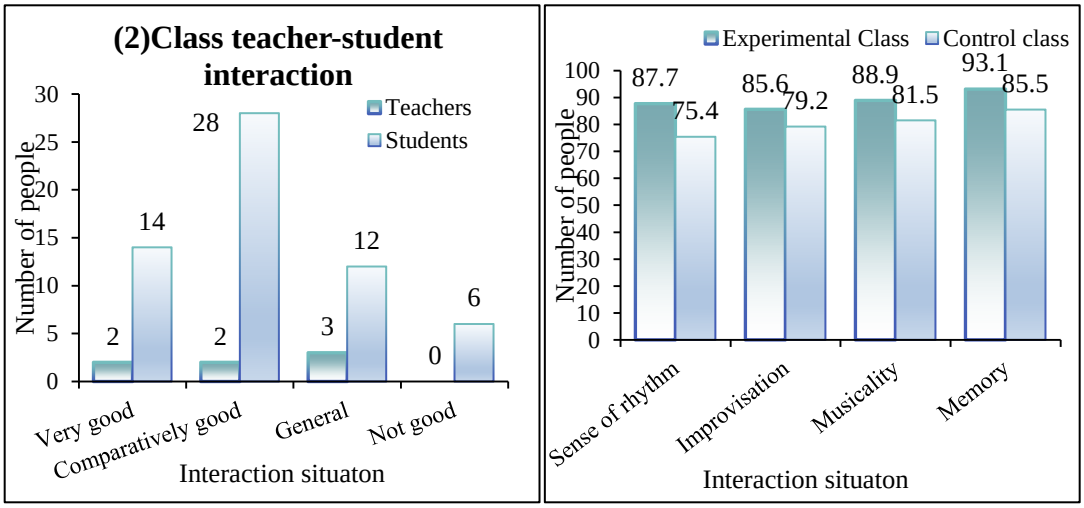

Figure 9. Teaching test results for two classes

It can be seen from Figure 9 that most of the indicators of the experimental class are better than those of the control class, and the average scores of the experimental class are all more than 85.5. Judging from the improvement of the ability to sense rhythm, most of the students in the experimental class can accurately strike the rhythm according to the more complex music melody, and can also combine different rhythm patterns to improvise. Practice has once again proved that the multimodal music teaching model has had a significant impact on improving the academic performance, music skills, sense of music, improvisation and memory ability of students in the experimental class.

In music teaching classes, teachers often use multimodal resources, and the multimodal resources used are very rich and balanced, including text, pictures, video, audio, etc. Among them, the frequency of use of text and pictures is as high as 90%. In the classroom, teachers use ppt courseware for teaching, and also use other forms of resources, such as body language, among which the use rate of gestures reached 65%, as can be seen in Figure 10:

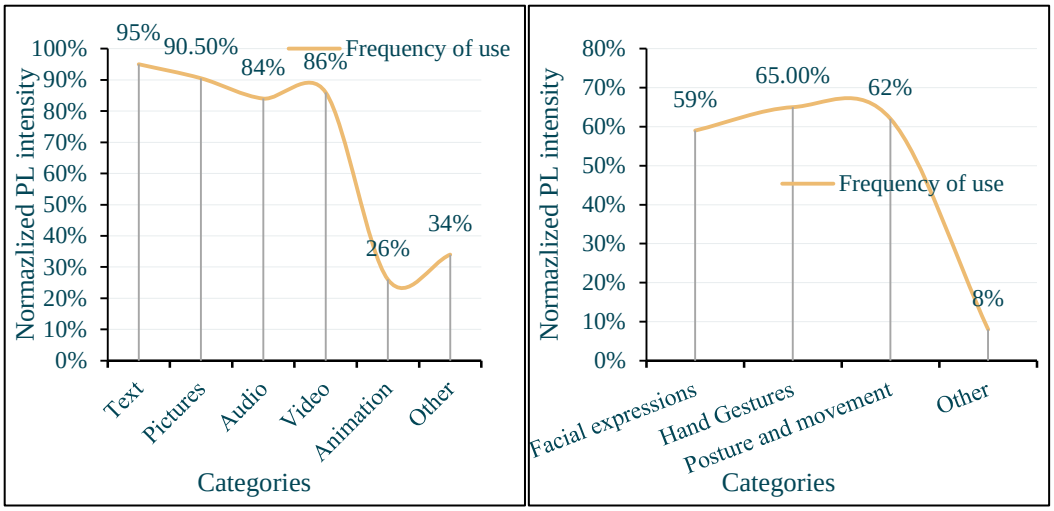

Figure 10. Multimodal resources in music classroom teaching

5. Conclusion

With the rapid development of multimedia and computer technology, traditional learning methods can no longer meet the needs of learning seekers, and new educational methods based on information technology are also coming out. The multi-modal teaching mode broke the previous single-modal teaching mode, greatly promoted students' classroom participation, and effectively promoted classroom learning. Under the conditions of informatization, the application of various forms of teaching modes in the classroom is inevitable. The multi-modal teaching mode created a real cultural situation by using modal resource information such as pictures, audio and video, and teachers' body language, and the coordination of multiple modalities can fully stimulate students' multi-sensory and deepen their understanding of the traditional learning. The memory and understanding of cultural content created a relaxed and pleasant teaching atmosphere, which made students interested and motivated in learning. The multimodal teaching mode can improve the classroom efficiency of music teaching to a certain extent. It is believed that there will be more and more practical researches on the application of multimodal teaching mode to various classrooms in the future.

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] *Fortuna S. Embodiment, sound and visualization: A multimodal perspective in music education. Zbornik Akademije umetnosti, 2017, 2017(5):120-131.*

[2] *Huang R. A Study on Multimodal Teaching Mode for Adult English Learners under Epidemic Situation. Open Journal of Modern Linguistics, 2020, 10(4):366-374.*

[3] *Lin W. Design research of english multimodal teaching system based on "internet". Revista de la Facultad de Ingenieria, 2017, 32(13):406-411.*

[4] *Shustrova E V., Checkletsova E I., Volkova L E. Teaching English by means of comic multimodal texts. XLinguae, 2017, 10(4):34-47.*

[5] *Yu F X., Suo Y N. Application of gesture recognition based on the somatosensory kinect sensor in human-computer interaction framework. Revista de la Facultad de Ingenieria, 2017, 32(10):580-585.*

[6] *Liang W. Scene art design based on human-computer interaction and multimedia information system: an interactive perspective. Multimedia Tools and Applications, 2019, 78(4):4767-4785.*

[7] *GAO, Yongqiang, LU. Vision-Based Hand Gesture Recognition for Human-Computer Interaction——A Survey. Wuhan University Journal of Natural Sciences, 2020, v.25; No. 130(02):82-97.*

[8] *Chen, Minghui. Multimodal Analysis in Academic Settings: From Research to Teaching. VISUAL COMMUNICATION, 2018, 17(4):475-478.*

[9] *Liu H., Jiang Y. Study on the multimodal teaching exploration of English and American literature course based on the internet plus background. Boletin Tecnico/Technical Bulletin, 2017, 55(19):110-116.*

[10] *Huang S Y. A critical multimodal framework for reading and analyzing pedagogical materials. English Teaching: Practice and Critique, 2019, 18(1):52-69.*

[11] *Kong W. Digital media art design based on human-computer interaction technology in the background of big data. Revista de la Facultad de Ingenieria, 2017, 32(14):485-489.*

[12] *Jing Y. Research on fuzzy English automatic recognition and human-computer interaction based on machine learning. Journal of Intelligent and Fuzzy Systems, 2020, 39(4):5809-5819.*

[13] *Hamond L., Himonides E., Welch G. The nature of feedback in higher education studio-based piano learning and teaching with the use of digital technology1. Journal of Music Technology and Education, 2020, 13(1):33-56.*

[14] *Jing C. Application and research on computer aided technology in the new music teaching method. IPPTA: Quarterly Journal of Indian Pulp and Paper Technical Association, 2018, 30(7):301-306.*

[15] *Arc S. Computer-Assisted Musicteachingin Music Teacher Education Departments: Marmara University Sample. Educational Policy Analysis and Strategic Research, 2018, 13(1):71-80.*

[16] *Moore D. Supporting students in music technology higher education to learn computer programming. Journal of Music, 2017, 7(1):75-92(18).*

[17] *Li F. Study on the interactive creation of computer assisted music teaching. Boletin Tecnico/Technical Bulletin, 2017, 55(14):685-691.*

[18] *Jin C. Analysis of music teaching mode innovation based on intelligent classroom and multimedia system. Revista De La Facultad De Ingenieria, 2017, 32(1):534-543.*

[19] *Chi X. Study on vocal music teaching innovation mode based on computer simulation and voice spectrogram analysis. Revista de la Facultad de Ingenieria, 2017, 32(16):400-406.*

[20] *Jing Y., Guo S., Wu X. Vocal music teaching mode based on a computer platform. Agro Food Industry Hi Tech, 2017, 28(1):2200-2204.*

[21] Kuang H. Research on Multimedia Music Teaching Based on Modern Multimedia Equipment. Revista De La Facultad De Ingenieria, 2017, 32(3):166-172.