

Research Article

Research on the Whole Teaching of Vocal Music Course in University Music Performance Major Based on Multimedia Technology

XiaoFang Zheng 

Xiamen University, Xiamen, Fujian 361000, China

Correspondence should be addressed to XiaoFang Zheng; dora3000@163.com

Received 18 December 2021; Revised 5 January 2022; Accepted 6 January 2022; Published 2 February 2022

Academic Editor: Maozhen Li

Copyright © 2022 XiaoFang Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

MT (multimedia technology) music teaching has changed the rigid, unitary, and boring teaching form of traditional music teaching, which has made a comprehensive breakthrough and impact on traditional university music education and has shown a brand-new teaching form to contemporary college students, greatly increasing their strong interest in learning music. In this paper, an objective evaluation method of vocal music quality based on the comparison of sound parameter characteristics is proposed by using the technology of sound signal parameter analysis and extraction. Through confidence measure, reliable data are selected for online adaptive updating of Gaussian mixture model, and the recognition results are smoothed to further remove the instantaneous mutation error. The main melody of singing is discriminated by fundamental frequency discrimination model. Experiments show that the overall accuracy rate of the main melody of vocal music extracted by this algorithm is 86.24%.

1. Introduction

MT (multimedia technology) is the product of the progress of the times and the demand of educational development [1]. With the development of the times and the progress of science and technology, MT and its related teaching methods have gradually played an increasingly important role in China's higher education system because of their synchronization, integration, and interactivity [2]. Vocal music teachers usually have professional knowledge of music theory but lack systematic computer knowledge. The evaluation of students' singing level is mainly based on their personal experience and feelings, with strong subjectivity. Therefore, it is of great practical and promotional value to study and develop an intuitive and easy-to-operate interactive multimedia vocal music teaching system.

The progression of music art from early concrete music to modern synthetic electronic music to post-modern computer music reflects not only the evolution of the music art form, but also the evolution of music technology [3, 4]. Music education at the university level is an important part

of modern education. It not only enriches young students' after-school activities, but also provides them with another professional skill, allowing them to enter society with a skill [5]. The advantages of MT include an intuitive image and strong interaction. Exploring and applying the benefits of MT can help students form scientific sound concepts, improve their intonation and rhythm, cultivate their expressive force in vocal music singing, and create favorable conditions for the formation of their personalized performance style [6, 7]. College vocal music teachers should gain a scientific understanding of the value of MT, find the best combination of technology and vocal music teaching, promote innovation and reform in the vocal music teaching mode, and improve the efficiency and quality of vocal music instruction.

With the continuous development of network technology and information technology, many universities have adopted MT in the course content of auxiliary harmony, solfeggio, impromptu accompaniment, etc., which can show the vocal music teaching process more intuitively, make the teaching content illustrated with pictures, and thus stimulate students' enthusiasm for learning vocal music [8]. In this

paper, an objective evaluation method of vocal music quality based on the comparison of sound parameter characteristics is proposed by using the technology of sound signal parameter analysis and extraction. Select reliable data through confidence measure to realize online adaptive updating of GMM (Gaussian mixture model), and get GMM that better matches the audio signal to be segmented, thus improving the recognition accuracy. In the song melody positioning part, based on the difference of timbre between song and accompaniment, the fundamental frequency discrimination model using neural network is added, and the probability that the dominant fundamental frequency track belongs to song melody based on subsection statistics can effectively reduce the false alarm rate of melody positioning and improve the overall accuracy.

2. Related Work

Literature [9] discusses three aspects: the historical evolution of music communication media, the appearance of new audio-visual media competition, and the dominant trend of network music communication. Literature [10] gives appropriate suggestions on the function and selection of modern teaching media. In [11] by comparing with the old standards, this paper analyzes the importance of the requirements of teachers' information and communication technology and cooperation ability in the professional competence standards for primary and secondary school teachers issued by the French Ministry of Education. In [12] by allowing students to actively participate in various music teaching practice activities and respecting their different learning methods and different music experiences formed as an independent individual, students' aesthetic ability can be comprehensively improved, so that students' creative thinking can be fully developed, thus forming their good humanistic quality. Literature [13] regards everyone as an independent individual, encourages students to dare to express their opinions, and encourages students to think creatively. On the basis of the existing theoretical analysis, this paper comprehensively analyzes the principles, learning modes, and implementation methods in the integration of information technology and music curriculum. In [14] according to the current situation that MT is applied to music teaching, through the breadth and depth of multimedia application and teaching results, this paper makes some thoughts and studies on the role and prospect of multimedia in the actual process of music education. Literature [15] points out that the correct use of multimedia teaching means in music teaching is of great significance to the improvement of music teaching theory and norms. Literature [16] studies people's activities in music education from the psychological point of view, including the psychological activities of students in feeling music, expressing music, creating music, learning music knowledge and skills and music culture, and psychological activities in music appreciation, etc., which are found out to guide educational activities. This paper is of great help to the research of music classroom in primary and secondary schools.

Speech plays an important role in modern life as a means of communication. People's demand for speech signal processing technology is growing every day as society and science and technology progress, and it has been vigorously developed, including speech coding, speech decoding, speech synthesis, speech recognition, speech enhancement, and so on. The AR (autoregressive) model is used in [17] to build a model of source signal separation that realizes the separation of single channel source signals. Pitch estimation is used to generate sound music templates in [18]. In [19] the amplitude spectrum of mixed signals is separated into a sparse matrix and a low-rank matrix using the sparsity of speech signals and the low-rank nature of music accompaniment, and then the separation of vocal music is realized using binary templates. The separation of different music signals is extended using nonnegative matrix decomposition. One of the major drawbacks of this method is that it requires a lot of computation and the separated signals are too simple. For nonnegative matrix decomposition, there are also some improved methods and supervision methods. A main melody extraction method based on speech separation has been proposed in [20, 21]. Experiments show that main melody extraction algorithms that include speech separation perform better than those that do not. Literature [22] calculates the saliency function using the source-filter model and harmonic weighting, extracts several candidate melody lines based on melody continuity, and finally locates the song melody and extracts the main melody based on a set of candidate melody line characteristics.

3. Research Method

3.1. Feasibility of Modern MT in Vocal Music Teaching. The ultimate goal of MT in music class is to increase the vitality of college music teaching activities. The multimedia class should be elegant, simple, and convenient to operate, the selection of teaching materials should be scientific and reasonable, other contents that are not helpful to the teaching focus should be stopped, and music content that is helpful but difficult for college students to understand should also be carefully selected.

MT can provide more abundant music information resources for autonomous learning for college students with high music literacy than in the classroom, and MT can provide key technical support for college students to carry out music extracurricular activities and online retrieval, and this type of autonomous learning in spare time is simply an effective extension of college music classroom learning. Students can choose the learning content that best suits them if they are made aware of autonomous learning. Furthermore, multimedia can assist students in learning independently by utilizing its inherent benefits and by providing more comprehensive learning resources. Using MT for music education is a good combination of music instruction by teachers in the classroom and independent learning by students outside of the classroom, and it gradually evolves into a new teaching mode that not only broadens students' horizons but also improves their learning efficiency.

Using MT for music teaching has become an inevitable choice for the change of teaching form, but multimedia is only a tool, which cannot replace teachers' position in the classroom. Therefore, it is necessary to strengthen the training of music teachers' ability to use MT and deepen teachers' understanding of MT and its significance to today's music teaching, not only to let teachers learn how to use MT, but also to make teachers learn how to use MT reasonably, so as to make it a tool to promote music classroom teaching rather than a tool to hinder music classroom teaching process.

Music aesthetic psychology refers to the psychological state and ability of the appreciator in aesthetic activities, including music perception, music thinking, and music emotion. The aesthetic performance is the difference of the appreciator's aesthetic attitude, aesthetic preference, and aesthetic ability. In music classroom teaching, apart from the differences in individual physiological and psychological performance, there are differences in their musical aesthetic psychology. Through the sound level, intensity, and rhythm, intuitively feel the brightness of the object, the ups and downs of modality, and the excitement of feeling. In hearing, the higher the pitch, the higher the sense of space, and the brighter the vision. In a good mood, the individual's active thinking and concentration make students more sensitive to the feeling and discrimination of the basic elements of music.

Multimedia teaching can transform a dull classroom into three-dimensional content, which can pique students' interest in learning by incorporating the vivid design of hearing, vision, and feeling appreciation that textbooks demand. Teachers cannot show concrete objects in areas where teaching conditions are limited, students do not recognize the sound effects of musical instruments in textbooks and pictures, and they are unfamiliar with the customs of a few areas. These difficult multimedia can be presented one at a time to pique students' interest and pique their curiosity, allowing them to appreciate and comprehend more stereoscopically.

From the age of seven to eleven years, they are called the early school age. They are in the specific operation stage, and their unintentional memory is dominant, their intentional memory is gradually strengthened, their memory of intuitive images is stronger than that of abstract logic, and the time of continuous concentration of attention increases with the development of age. The research on memory ratio, as shown in Figure 1, also learns a content. After comparison, it is found that through the combination of audio and video, people can master knowledge more firmly and more efficiently through various senses.

In the stage of formal operation, intentional memory is dominant in middle school students. In attention, they can take the initiative to pay attention according to their own preferences and purposes. With the increase of age, their ability to understand music and control their bodies is enhanced, and their learning efficiency is naturally enhanced. When the knowledge learned is logically arranged across the length of time and presented to the students, the students can absorb it at a glance. The collection and editing of audio-visual

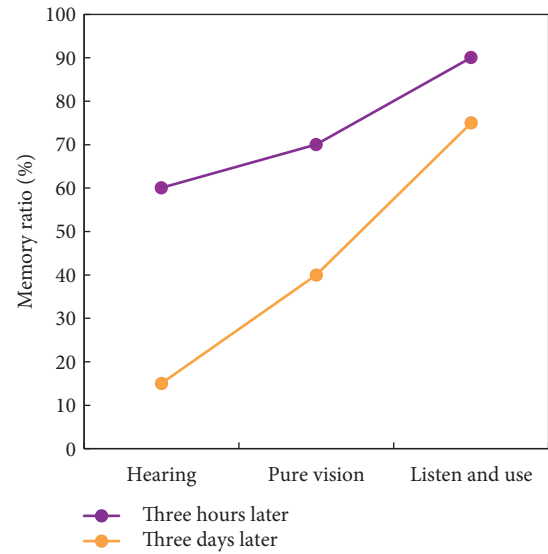


FIGURE 1: Memory ratio.

materials can enable the teachers to extract the important and difficult points independently and save the time and efficiency of repeated appreciation from the beginning to carry out other teaching activities and improve the teaching efficiency.

Multimedia teaching can provide a good atmosphere for students. Sensory stimulation can make students' music imagination and music thinking active. This environment enables students to follow the teacher's steps. They can also have an epiphany during this period, which is also the sublimation stage of music aesthetics, enabling them to complete the teaching content efficiently.

If teachers make rational use of differences in college students' knowledge structures, carefully set teaching objectives, and create scientific teaching design, it will not only help to promote college students' music learning, but also help college students learn as much about music knowledge as possible in a short class by taking advantage of the cross between disciplines and deepen their understanding of music and related art culture, in order to realize the educative goal.

3.2. MT-Based Holistic Teaching Method of Vocal Music Course for College Music Performance Major

3.2.1. Vocal Music Evaluation Method. The key to the research and development of multimedia vocal music teaching system is to establish the corresponding vocal music measurement method and scoring mechanism. Vocal music scoring is different from voice scoring. The multimedia vocal music teaching system established in this paper adopts the scoring mechanism of standard sound materials [23, 24], and the vocal music singing quality evaluation system using standard sound materials mainly consists of three parts: sound feature extraction, feature parameter matching, and scoring mechanism, as shown in Figure 2.

The corresponding feature parameters are extracted and matched after preprocessing the evaluated singing

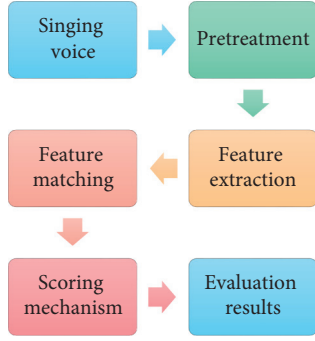


FIGURE 2: Vocal music evaluation system based on feature comparison.

voice and a standard voice, respectively. The higher the similarity, the higher the score given by the scoring mechanism to the evaluated voice. The volume intensity curve, fundamental frequency trajectory, and breath smoothness are the most important feature extraction parameters. The dynamic time warping method is used to compare similarity in feature matching.

Singing refers to performing activities with human body as sound generator. When singing, human body and spirit should enter and maintain the required state of singing. Correct breathing is the driving force of singing. Therefore, breath plays an important role in singing. The measurement of breath is a process of self-comparison, and the degree of breath stability can be measured by calculating the standard deviation of test sound waveform.

Standard deviation is a measure of the degree of dispersion of a set of values from the mean value. A larger standard deviation means that most of the values differ greatly from their average values, while a smaller standard deviation means that most of the values are closer to the average values. For vector X , its standard variance function $\text{std}(X)$ is

$$\text{std}(X) = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}, \quad (1)$$

where n represents the number of sampling points; \bar{x} represents the average amplitude.

In sight singing, syllables and notes have a one-to-one correspondence. Seven singing names have no zero initial consonant, so the vowels are alternated. The score can be used to determine the roll call of notes, and phonetics knowledge can be used to predict the acoustic characteristics

of each phoneme. As a result, note segmentation can be transformed into the following problems if the singer does not omit or add notes:

Given the score sequence M and the singing sequence S , solving an optimal match T between them can minimize the cost function, that is:

$$\arg \min \text{Cost}(T|S, M). \quad (2)$$

Research shows that people's perception of pitch depends on the pitch of the most stable part inside the note. Therefore, the histogram of all sampling points on the pitch curve is taken according to the preset granularity, and the component with the most sampling points in the histogram is taken as the actual pitch of the note. On this basis, the pitch deviation between singing S and reference score M is defined as the average value of the difference between the actual pitch of all notes and the pitch of reference score (after transposition) weighted by the length of notes, namely:

$$\text{pitch_diff} = \frac{\sum_{i=1}^N \text{Dur}_i \times |\text{pitch}M_i - \text{pitch}S_i|}{\sum_{i=1}^N \text{Dur}_i}. \quad (3)$$

$\text{pitch}M_i$ represents the pitch of the i th note in M , $\text{pitch}S_i$ represents the actual pitch of the i th note in S , Dur_i is the duration of the i th note, and N is the total number of notes in the song.

Duration accuracy reflects the consistency of note duration in singing. Similar to rhythm, the score does not specify the absolute length of notes, so it is necessary to normalize the length of notes in actual singing by using the beat number of notes in the score. In the concrete implementation, this paper uses the standard deviation coefficient of normalized note duration to measure the duration deviation between score M and singing S , namely:

$$\text{dur_dev} = \frac{\text{std}(\text{dur}S_i/\text{dur}M_i)}{\text{mean}(\text{dur}M_i)}, \quad i \in [1, N], \quad (4)$$

where $\text{dur}M_i$, $\text{dur}S_i$ represent the duration of note i in M and S , respectively; $\text{mean}(A)$, $\text{std}(B)$ represent the mean and standard deviation of A , B .

The purpose of vocal music score is to give an objective evaluation of the singer's grasp of the melody of the music, mainly including sound intensity, pitch, and breath. The higher the score, the more accurate the singer's interpretation of the music; the lower the score, the less accurate the singer's grasp of the melody of the music. The scoring formula is as follows:

$$\text{score} = k_1 - \frac{100}{1 + a_1 (\text{min disp})^{b_1}} + k_2 \frac{100}{1 + a_2 (\text{min disp})^{b_2}} + k_3 \frac{100}{1 + a_3 (\text{st})^{b_3}}. \quad (5)$$

Among them, $k_1 \sim k_3$ is the weight of each scoring parameter in the scoring mechanism; min disv, min disp are the distance of sound intensity and pitch parameter, respectively; st represents the breath stability parameter. The choice of weights can be adjusted according to different requirements or different scoring priorities.

3.2.2. Extraction of Main Melody of Vocal Music. Melody is the most important musical element, which is composed of single tones with different pitches and durations. The main melody can be divided into vocal melody and general melody. If polyphonic music contains songs, the pitch sequence of songs is considered as the main melody of vocal music. If there is no singing voice, the pitch sequence of the instrumental playing sound dominated by energy is the main melody of the instrumental music.

From the perspective of physics, pitch is determined by fundamental frequency. Based on the characteristics of singing voice, this paper proposes an automatic labeling model (as shown in Figure 3).

Short-term stationarity is a property of music, which means that the signal characteristics are essentially stable over a short period of time. Stationarity can last anywhere from a few hundred milliseconds to several seconds depending on the characteristics of the music signal. The continuous nonstationary signal stream is segmented into a series of short-term stationarity segments, which is known as segment presegmentation.

Segment presegmentation divides the music signal into many segments s_i , $i = 1, 2, \dots, N$, and assumes that each s_i belongs to either vocal segment or nonvocal segment.

Let $|x_{i,1}, x_{i,2}, \dots, x_{i,M}|$ be the M -frame feature vector of s_i segment and let λ_v, λ_{nv} be the GMM (Gaussian mixture model) of vocal music and nonvocal music, respectively, and the logarithmic likelihood ratio of s_i to λ_v, λ_{nv} is as follows:

$$\begin{aligned} \log P(s_i | \lambda_v) &= \sum_{j=1}^M \log P(x_{i,j} | \lambda_v), \\ \log P(s_i | \lambda_{nv}) &= \sum_{j=1}^M \log P(x_{i,j} | \lambda_{nv}). \end{aligned} \quad (6)$$

If $\log P(s_i | \lambda_v) > \log P(s_i | \lambda_{nv})$, then s_i is identified as a vocal category; otherwise it is a nonvocal category.

One of the key technologies for music segmentation is the training of vocal and nonvocal music models. The accuracy of recognition is influenced by the quality of the models. The instruments, playing methods, singers' voices, and vocals are all very different because of the various types and genres of music and songs. As a result, the key technologies for improving segmentation accuracy are reducing model complexity and bringing them closer to each piece of music to be processed.

Amplitude compressed pitch estimation filter is a robust multifundamental frequency extraction method [9, 10]. In this paper, the pitch saliency function is calculated by convolution of logarithmic domain comb filter and logarithmic domain spectrum. A voiced signal has fundamental frequency f_0 , whose frequency domain expression is

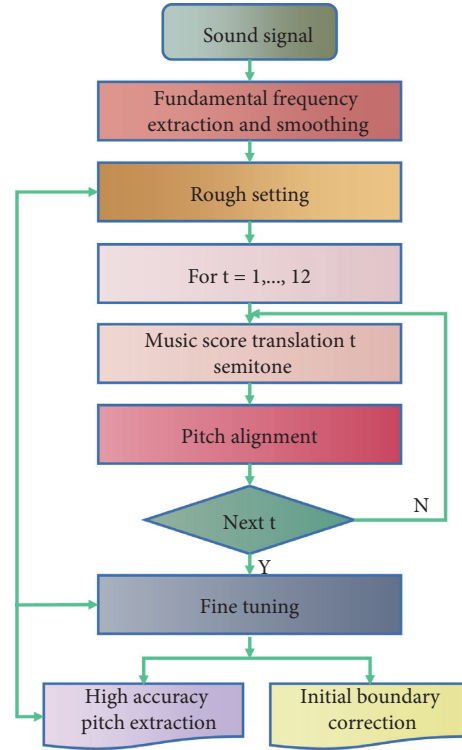


FIGURE 3: Automatic vocal music annotation model.

$$Y(f) = \sum_{k=1}^K a_k \delta(f - kf_0), \quad (7)$$

where a_k is the coefficient of the k harmonic.

Both singing and instrumental music have harmonic structure, so the mixed music spectrum has approximate sparsity. The comb filter can be used to extract the harmonic spectrum of the corresponding sound source according to the dominant fundamental frequency, and the MFCC (Mel frequency cepstral coefficients) of the extracted signal is sent to the neural network to judge whether the corresponding fundamental frequency is singing. Main melody discrimination steps:

A comb filter with a frequency range of 0–4 kHz is constructed from the dominant fundamental frequency F_0 , as shown in

$$h(f) = \sum_{k=1}^K \delta(f - kF_0) * b(f), \quad (8)$$

where K is the number of harmonics in the range of 0–4 kHz; $b(f)$ is the basic waveform of comb filter (rectangle is used in this paper).

The MFCC is sent to the neural network to identify whether the dominant fundamental frequency F_0 is the fundamental frequency of singing.

Count the frame number of the fundamental frequency of the voice in each voiced segment, and if it is more than half of the total frame number of the voiced segment, determine that the dominant fundamental frequency track of the voiced segment is the main melody of the voice.

4. Results Analysis and Discussion

4.1. Evaluation and Analysis of Vocal Music. In this paper, the vocal music quality measurement method and scoring system are simulated on Matlab. The scoring test is mainly aimed at skill training etudes, which are the most commonly used etudes in vocal singing training. Focus on specific vowels, phonetic syllables, and skilled vocalization for targeted training. In the experimental simulation, five basic vowel vocal training materials *A*, *E*, *I*, *O*, and *U* and male closed humming training songs are selected for testing and analysis.

In the comparison of sound intensity, the average amplitude of each frame signal is calculated by the standard of practice music and the audition music, respectively, as the sound intensity parameter of this frame, thus drawing the volume intensity curve, as shown in Figure 4.

Although MT can bring many benefits to music teaching, teachers should grasp the utilization of MT and cannot let it replace teachers in teaching music knowledge. Teachers use MT in music classroom to improve classroom efficiency and teaching vitality. Therefore, in courseware making, they should not add too much content that is not related to classroom teaching, but should ensure that the whole courseware is simple and generous, which is convenient for teachers to operate and students to watch.

In pitch comparison, the fundamental frequency tracks of two pieces of music are obtained separately by cepstrum method, as shown in Figure 5. It can be found that the average standard fundamental frequency is 157.36 Hz and the average audition fundamental frequency is 159.01 Hz.

When calculating similarity, DTW (Dynamic Time Warping) method is used to get the average distance of the closest two features. Figure 6 is the DTW comparison between the standard song and the test song.

Here, it is important to point out that no matter how powerful and convenient MT is, it is only a teaching aid, which can only help teachers to impart knowledge and skills. Therefore, in the practice of vocal music teaching, attention should be paid to teacher-student interaction and the cultivation of the comprehensive quality of teachers and students. MT brings an opportunity for vocal music teaching innovation, which requires teachers to seize the opportunity, actively reform the teaching mode, enliven the classroom teaching atmosphere, break the limitations of traditional vocal music teaching, and let students have more room for development and innovation.

Teachers can use MT to simulate the stage background and stage performance situation for students, so that students can actively participate in stage performance activities, show their talents and skills, and increase their artistic expression. For example, in the practice teaching of vocal music performance for students, in order to bring students into the corresponding performance situation and motivate students to perform and show actively, teachers can use MT to simulate the stage background and stage performance situation for students, so that students can actively participate in stage performance activities, show their talents and skills, and increase their artistic expression. When students arrive, they will have a better understanding of vocal music knowledge and skills. They will also have close communication and interaction with other students,

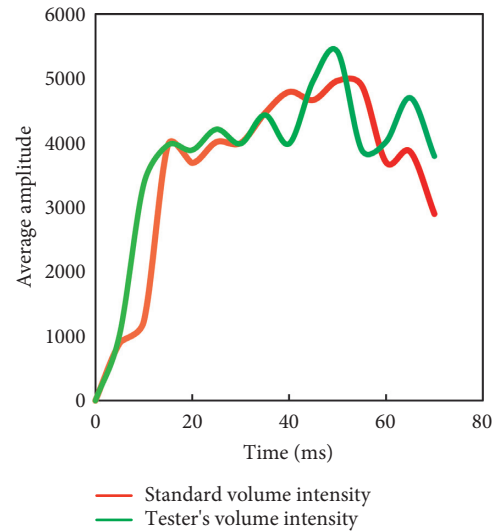


FIGURE 4: Men's voice closed hum practice music sound quantity intensity by line.

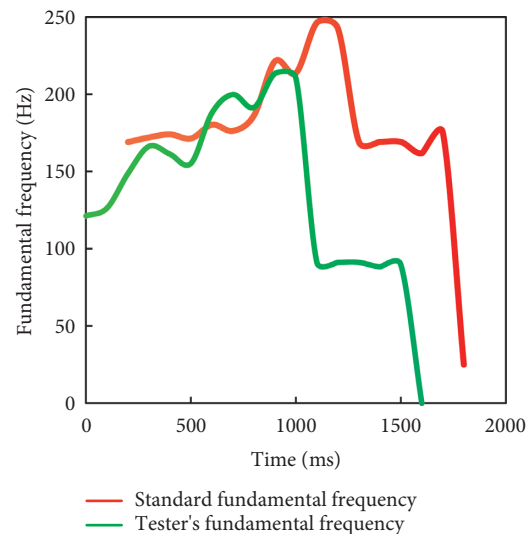


FIGURE 5: The track of the fundamental frequency of the male closed-mouth humming practice song.

learn some useful methods and skills, and save money for their future growth and progress.

4.2. Extraction and Analysis of Main Melody. Figure 7 shows the segmentation error rates of three models (initial model, model updated with all data, and model updated with reliable data) without smoothing the recognition results. For each model, the final segmentation error rates are compared when the presegmentation segment length changes from 0.1 to 1.0 s.

As can be seen from Figure 7, the data selection model updating algorithm based on confidence measure proposed in this study significantly reduces the error rate of music segmentation, and the segmentation result is further improved by smoothing the recognition result.

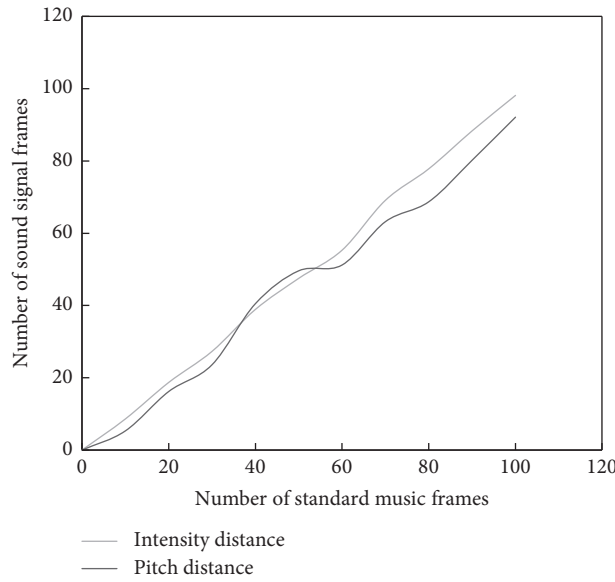


FIGURE 6: DTW comparison between standard song and test song.

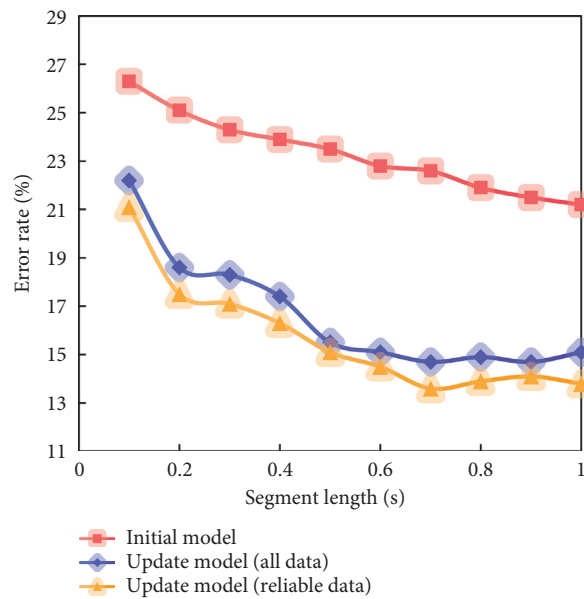


FIGURE 7: Unsmoothed processing.

Figure 8 shows the corresponding segmentation error rate after smoothing the recognition result, where the error rate is defined as the percentage of the length of the incorrectly classified music signal to the total signal length.

Figure 8 shows that the smoothing process is effective when the length of presegmented segments is short. With the increase of segment length, the smoothing process causes each continuous segment to be too long but introduces additional segmentation errors.

Therefore, in music teaching, the subject should always be grasped, use MT appropriately and moderately, and do not become the teaching slave of MT. The use of modern

educational technology means such as multimedia is just one of the ways of expressing teachers' creativity, a kind of teaching means and teaching methods, and a tool for serving teachers. To use it reasonably, it should not be used lightly; of course, it should not be used greatly, but it should be used by me. Teachers cannot rely entirely on MT to reflect their teaching level.

Figure 9 shows the minimum error rate value on each curve in Figures 7 and 8, that is, the best result obtained when the presegmentation length of each algorithm changes.

It can be seen from Figure 9 that the segmentation error rate is reduced from 18.6% to 13.9%. Compared with the

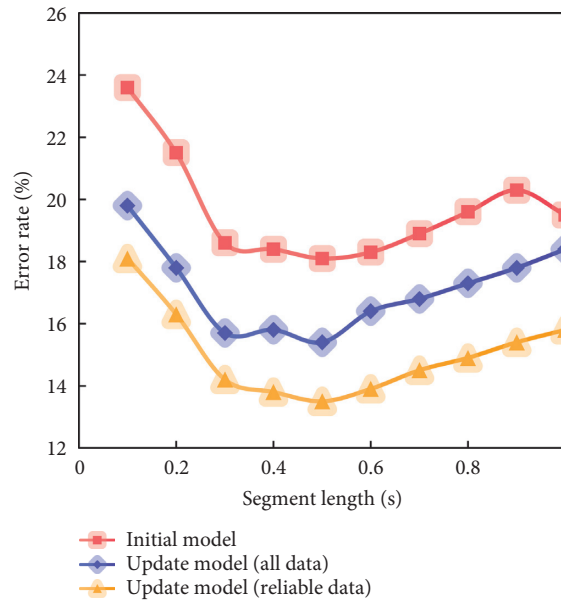


FIGURE 8: After smoothing.

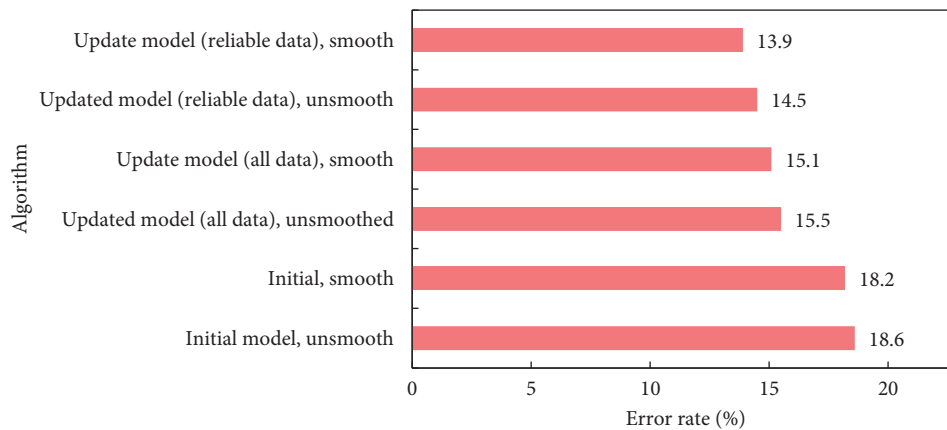


FIGURE 9: Comparison of music segmentation error rate.

original model and without smoothing, the segmentation algorithm proposed in this study has improved the error rate.

In order to verify the accuracy of the automatic extraction algorithm of vocal music melody, this section uses 500 pieces of music in the test set as test data and carries out the experiment under the condition that the signal-to-interference ratio is 5 dB, respectively. The experimental results are shown in Figure 10. Among them, the following five performance indexes are used to evaluate the algorithm performance [19]: VRR (Voicing Recall Rate), VFAR (Voicing False Alarm Rate), RPA (Raw Pitch Accuracy), RCA (Raw Chroma Accuracy), and OA (Overall Accuracy).

It can be seen that the overall accuracy of the main melody extracted by this algorithm reaches 86.24%. In this paper, the fundamental frequency discrimination model is

introduced, and the statistical method is used to judge whether the dominant fundamental frequency track of each voiced segment belongs to the main melody of the song. Therefore, in rare cases, the song melody segment will be misjudged as the accompaniment melody segment, which will reduce the recall rate of melody location. However, in most cases, the accompaniment melody segment will not be judged as the song melody segment, which will reduce the false alarm rate of melody location and help to improve the overall accuracy rate of the algorithm.

To improve the comprehensiveness of students' knowledge in college vocal music teaching, in addition to thoroughly learning the teaching materials, it is necessary to increase classroom teaching capacity and expand teaching information with MT, so that students can fully grasp the connotation of works in future vocal music learning and

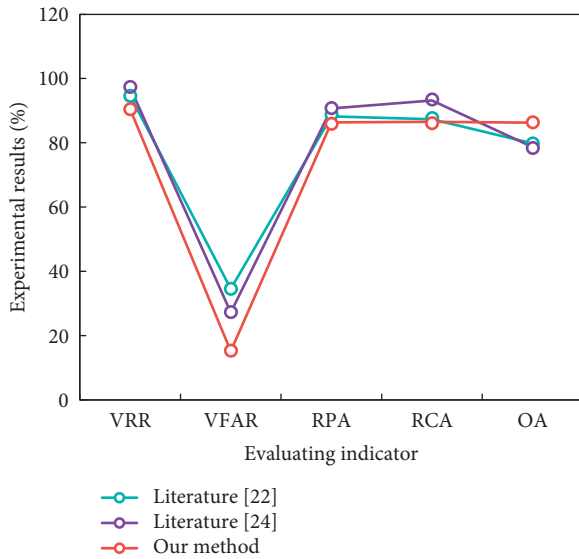


FIGURE 10: Experimental results when the signal-to-interference ratio is 5 dB.

practice. In the past, students could not form a systematic and deep impression from a single classroom explanation, nor could they access as much knowledge and information as possible. As a result, teachers can introduce MT, make full use of computers' powerful storage capabilities, actively consult a variety of materials and information, and provide rich information resources for vocal music instruction.

5. Conclusion

Combining college music teaching with MT meets the needs of the times. It can not only increase students' learning desire and improve students' learning ability, but also have considerable influence on teachers' teaching quality. It is the art of music, the art of practice, and the art of emotion. Simply emphasizing the description of language is boring for students. Teachers should understand students' aesthetic psychology of music when designing classes. By analyzing and displaying the waveform of the singer's voice, this paper makes vocal music teaching intuitive and visible. On this basis, according to the short-term stationary characteristics of the music signal, the instantaneous mutation error of the segment in the recognition result is further removed by smoothing the recognition result. Experiments show that the main melody extraction algorithm in this paper can effectively reduce octave error with a signal-to-interference ratio of 5 dB, the false alarm rate of melody location is obviously lower than other algorithms, and the overall accuracy is higher than other algorithms, which can effectively extract the main melody of vocal music.

However, there are still many shortcomings in MT teaching. Teachers need to improve the problems step by step according to their own actual teaching situation and improve the teaching quality to a new field by updating MT, so as to contribute to the cultivation of high-quality music talents in China.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] S. Wang, "Research based on multimedia technology in college vocal music classrooms," *Scenic Spots*, vol. 7, no. 7, p. 1, 2019.
- [2] Y. Li, "Research on vocal music teaching in colleges and universities based on multimedia technology," *Shenhua: Shang*, vol. 11, no. 11, p. 2, 2019.
- [3] Y. Zhao, "Research on the overall teaching of the vocal music class of music performance in colleges and universities," *Science Fiction Pictorial*, vol. 11, no. 11, p. 2, 2018.
- [4] D. Chang, "Research on the application and function of multimedia technology in vocal music teaching in colleges and universities," *Literary Life-Art Theory*, vol. 227, no. 4, p. 229, 2019.
- [5] X. Zhang, "Research on the application and role of multimedia technology in vocal music teaching in colleges and universities," *Literary Life-Literary Theory*, vol. 3, no. 3, p. 226, 2018.
- [6] G. Zhang, "Discussion on vocal music teaching based on multimedia technology," *The North Music*, vol. 14, p. 128, 2021.
- [7] R. Wang, "The application and discussion of multimedia technology in music teaching," *Food Research and Development*, vol. 41, no. 23, p. 262, 2020.
- [8] W. Huang and B. Chen, "Vocal recital course reform and practice—based on the "research, performance, competition, and study" integration of "guide schema" research," *Journal of Jimei University*, vol. 20, no. 1, pp. 80–84, 2019.
- [9] Z. Guo, "Research on the application of internet and network multimedia in vocal music teaching," *Literary Youth*, vol. 19, no. 19, p. 1, 2019.
- [10] Y. Liu, "Research on the application of multimedia technology in traditional vocal music teaching," *The Voice of the Yellow River*, vol. 11, no. 11, p. 2, 2020.
- [11] Z. Cui, "The effective integration of vocal music teaching and information technology—comment on "music teaching and multimedia technology application in colleges and universities," *Chinese Science and Technology Paper*, vol. 14, no. 10, p. 1, 2019.
- [12] W. Deng, "Exploration of vocal music teaching based on multimedia technology," *Tomorrow Fashion*, vol. 10, no. 4, p. 1, 2018.
- [13] C. Yuan, "A brief analysis of the application and practice of multimedia technology in vocal music teaching," *Comedy World*, vol. 661, no. 11, pp. 54–55, 2020.
- [14] Q. Du, "Exploring the reform of vocal music teaching in higher vocational music majors under the mode of combining work with learning," *Chinese Journal of Multimedia and Network Teaching (Mid-term)*, vol. 12, no. 12, pp. 14–15, 2018.
- [15] L. Huang, "The application of multimedia technology in vocal music teaching," *Shaanxi Education: Higher Education Edition*, vol. 8, no. 8, p. 2, 2021.

- [16] R. Wang, "Discussion on the application of multimedia technology in the teaching of students' vocal music," *Satellite TV and Broadband Multimedia*, vol. 5, no. 5, p. 2, 2020.
- [17] G. J. Aguiar, R. G. Mantovani, S. M. Mastelini, A. C. P. F. L. de Carvalho, G. F. C. Campos, and S. B. Junior, "A meta-learning approach for selecting image segmentation algorithm - ScienceDirect," *Pattern Recognition Letters*, vol. 128, pp. 480–487, 2019.
- [18] L. R. Dung, Y. M. Yang, and Y. Y. Wu, "A hierarchical grab cut image segmentation algorithm," *Journal of Computer and Communications*, vol. 6, no. 2, p. 8, 2018.
- [19] M. T. Pearce, "Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation," *Annals of the New York Academy of Sciences*, vol. 1423, no. 1, pp. 378–395, 2018.
- [20] L. Bouaffif and N. Ellouze, "Speech-music-noise discrimination in sound indexing of multimedia documents," *Sound and Vibration*, vol. 52, no. 6, p. 9, 2018.
- [21] T. Popescu, R. Widdess, and M. Rohrmeier, "Western listeners detect boundary hierarchy in Indian music: a segmentation study," *Scientific Reports*, vol. 11, no. 1, p. 3112, 2021.
- [22] H. M. Mallette, W. George, and I. Blum, "Segmenting the audience attending a military music festival," *International Journal of Event and Festival Management*, vol. 9, no. 1, 2018.
- [23] D. Tatsuya, "Neurophysiological markers of statistical learning in music and language: hierarchy, entropy, and uncertainty," *Brain ences*, vol. 8, no. 6, p. 114, 2018.
- [24] Y. Xu and S. Xu, "A clustering analysis method for massive music data," *Modern electronic technology*, vol. 5, no. 1, p. 8, 2021.