Retraction

Retracted: Design and Application of Vocal Music Teaching Management System

Mathematical Problems in Engineering

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Sudu, "Design and Application of Vocal Music Teaching Management System," *Mathematical Problems in Engineering*, vol. 2022, Article ID 5023333, 8 pages, 2022.

Research Article

Design and Application of Vocal Music Teaching Management System

Langteng Sudu [ID]

Alxa Vocational &Technical College, Alxa, Inner Mongolia 750300, China

Correspondence should be addressed to Langteng Sudu; 1219010229@st.usst.edu.cn

This article proposes a virtual reality-based management system for vocal music instruction. Additionally, this article proposes an improved algorithm for automatic vocal main melody extraction. The pitch saliency function calculation method is improved based on the spectral characteristics of vocal music signals in order to reduce the computational complexity and time required to extract the vocal main melody. The model presented in this article has the potential to increase the recognition accuracy of the main melody model, decrease the rate of melody localization false alarms, and increase the overall accuracy of vocal main melody extraction. Additionally, this article incorporates the extraction algorithm into the management system, making it convenient for teachers to use during instruction.

1. Introduction

The advancement of society has prompted the development of new requirements for university education. Universities should not only complete the cultivation of students' professional literacy but also the cultivation of students' comprehensive humanistic literacy and endeavor to promote the multidimensional development of high-quality talents for the benefit of society. As an important component of aesthetic education as well as humanistic education, vocal music education is a topic that colleges and universities are strongly concerned about in this [1, 2] context. Vocal music education courses have been developed at a number of universities in the hopes of improving students' overall education through the medium of vocal performance. However, the current good role of vocal music education in realizing the comprehensive educational function of university teaching has not been adequately exploited in terms of its current positive impact. This is mostly due to the fact that colleges and universities typically fail to develop a comprehensive and profound grasp of the teaching management model of vocal music education, which ultimately results in the growth of vocal music education as a discipline. There are numerous instructional management issues that arise in the real world. In order to achieve this, universities should devote more attention to the development of a scientific teaching management model for vocal music education, address the issues surrounding teaching management, and implement effective measures to raise the level of teaching management in vocal music education. It was decided to develop the vocal music instruction management system [3–6].

Teaching management systems have become increasingly important as the popularity of the OBE education concept has grown, as it has the popularity of engineering education certification in colleges and universities. The traditional teaching management system is no longer capable of meeting the needs of talent development in the modern-day. The teaching of vocal music is distinct from the teaching of traditional subjects. First and foremost, the practical instruction of vocal music is constrained by time and space constraints, making it impossible to meet the particular needs of students [7]. In 2020, the new crown virus is ravaging the world, and colleges and universities are delivering theoretical instruction online, but because of time and space constraints, they are unable to do practical vocal music instruction in person. Despite the fact that students

can practice at home, the effect of practice cannot be determined in a timely and precise manner [8].

Due to the high level of practicality associated with vocal music instruction, students gain an incredible amount of emotional experience and expression during the singing process, which is a factor that is rarely considered in the current field of virtual reality education. Regrettably, a number of issues plague some of the few virtual reality teaching applications available. These include a lack of authenticity, an unpleasant interactive experience, a visual instruction with limited visual content, and a lack of fluency. The following are the most critical points to bear in mind regarding these concerns [9, 10].

According to the first point, some researchers are utilizing virtual panoramic video technology to generate three-dimensional panoramic image data, which allows singers to follow the preset teaching mainline in order to conduct audio-visual interaction within the panoramic range and obtain a variety of different styles of audio-visual experience effects. Perception experiences are limited by a lack of in-depth interactive information, a lack of vivid display of singing emotions, and facial expressions, as well as the difficulty of establishing an interactive function at a deep level.

Another method is to create depth information of a three-dimensional scene using next-generation engine technology, which can provide users with a more comprehensive selection of scenarios for singing interaction, as well as certain facial expression binding, but the rendering and smoothness are not as good as they could be. Authenticity is lacking in this game, the graphics algorithm is not optimized enough, and the real-time calculation of the screen is susceptible to latencies [11–13]. Also, there are some state-of-the-art methods, such as clustering-guided particle swarm feature selection algorithm for high-dimensional imbalanced data with missing values and feature selection method, but they cannot handle music melody feature extraction well [14–16].

Following an assessment of the benefits and drawbacks of the aforementioned technologies, this research seeks to employ virtual reality engine technology as the primary production tool in order to conduct a systematic optimization design for modeling, facial expression, gesture capture, and multiangle camera recording and to apply this to vocal singing. More realistic, delicate, interactive, and diverse application methods are provided in the research of the thesis. Additionally, this article examines the issues that exist in the current management mode of college vocal music education and teaching and makes appropriate recommendations for resolving these issues. Finally, this research introduces and implements note segmentation and basic frequency discrimination models. By combining note-based time persistence with spectral distance-based note segmentation, the song is divided into paragraphs with a relatively constant frequency spectrum, which is advantageous for melody tracking and singing melody location during the performance. In the singing melody localization section, an additional fundamental frequency discrimination model based on a neural network is used, and the probability that the dominant fundamental frequency trajectory belongs to the singing melody is calculated using segmentation statistics. This can significantly reduce the melody localization's false alarm rate and improve overall accuracy.

2. Problems and Countermeasures in the Teaching Management Mode of Vocal Music Education

2.1. Analysis of the Problems Existing in the Current Management Model of College Vocal Music Education

2.1.1. Backward Teaching Method. The vocal music major has progressively arisen and developed as a result of the constant enrichment of people's artistic lives, and it is particularly beloved and acknowledged by the vast majority of students. Vocal music majors have been created at many colleges and universities in recent years, in order to gain greater social recognition while also improving their own art education levels. This has resulted in a steady stream of vocal music talents for the society, but the reality is that this isn't necessarily the case. Many colleges and universities are experiencing a problem of blind expansion, and the teaching system and management are unable to keep up with the pace of professional opening, resulting in a variety of issues such as chaotic management and substandard teaching effects in the teaching of vocal music, which has a negative impact on the level of vocal music teaching and the quality of vocal music majors. Reputation and social approval are important factors. A issue exists when some comprehensive colleges and universities mindlessly replicate art schools when it comes to the curriculum of vocal music, and the curriculum is not acclimatized, which has an impact on the depth and quality of vocal music instruction. The establishment of vocal music majors in some colleges and universities was based solely on market demand and market conditions, with no systematic teaching resources allocated. As a result, the quality of vocal music education has deteriorated, and the society's demand for vocal music talents has not been met, and students have not been provided with a systematic and comprehensive system based on the actual situation of society. Vocal instruction: this issue will not only have an impact on the quality of teaching in the vocal music major, but it will also have an impact on the construction and development of the entire school [17].

2.1.2. Lack of Professional and Reliable Teachers. Professional level and aesthetic literacy of vocal music teachers are intimately related to the quality and level of vocal music professional teaching in general and vocal music in particular. It is vital to introduce high-level professional vocal music teachers and to establish a dependable vocal music teaching team in order to increase the overall effect of vocal music education. Many college instructors, on the other hand, lack the essential expertise and personal accomplishment, and the overall quality of their instruction is poor. Student interest and initiative in learning are affected as a result of their inability to provide students with a

positive learning environment and atmosphere. Students also struggle to master the necessary vocal skills and are unable to comprehend the depth of meaning contained within the various vocal works.

2.1.3. Lack of Integration of Practice and Innovation in Teaching.
First and foremost, the teaching style of the teacher is a significant issue. The conventional method of instruction has survived to the present day. In the classroom, only the teacher is responsible for instructing, while the students' only responsibility is to listen. With practice, there is no targeted mix that works. The two are integrated through the detailed division of education majors. In addition, there is a dearth of pragmatism in the classroom environment. An excellent song must be altered and artistically recreated in order for the listener to retain its memory value. Some colleges and universities, on the other hand, do not offer these courses. There are 56 nationalities represented by 56 phrases, and what belongs to a nation is also what belongs to the world. National vocal music is becoming increasingly popular as a result of the influence of Western music and modern music, and many artists are combining contemporary popular elements that are very well integrated into national vocal music, such as the folk singer Tan Jing, who, on the singer stage, has successfully integrated various ethnic musical instruments into modern music. These days, the pop king Jay Chou's songs "Chrysanthemum Terrace" and "Dongfeng Break" not only feature ethnic tones, but they also incorporate national musical instruments into their compositions. The marriage of identity and modernity necessitates high-quality creation, which is something that many educational courses fail to provide. Furthermore, the instruction is stale, and there is no room for innovation. When teachers teach professional skills, they continue to sing the songs from the original set. Students will not be interested in listening if they do not keep up with social trends. Students will not be interested in the professional field if this is the case. If you lose interest in a course, the educational impact of courseware is poor [18, 19].

2.2. Strategies to Improve Vocal Music Education

2.2.1. Carrying Out Teaching That Combines Theory and Practice.
The objective of vocal music education is to develop professional vocal music abilities for the benefit of society, in order to be able to produce higher-level vocal music compositions in the future. For this reason, it is vital to combine theory and practice in order to carry out vocal music teaching activities in a collaborative manner in order to significantly increase the overall quality of vocal music education. In the field of vocal music education, vocal skills serve as the foundation for the entire curriculum. The mastery of vocal abilities and the development of good fundamental skills allow pupils to create a solid foundation for the study of subsequent works as well as the expression of emotions. The development of students' theoretical knowledge should be a priority for teachers, who should help

students better grasp music theory knowledge and diverse music genres, as well as assist students in using theoretical information to improve their vocal music performance skills.

The teaching substance of vocal music theory is hard, and the learning process is generally monotonous, which can easily detract from students' interest in the subject matter being taught. Teaching methods and concepts such as multimedia teaching, simulation teaching, scene reproduction teaching, and other strategies can be used to increase students' interest in learning vocal music theory. Students must understand the role and importance of theoretical knowledge in order to achieve a good teaching effect during the course of theoretical learning in order to achieve good teaching results. Additionally, increase the number of class hours dedicated to students' practical activities so that students can fully appreciate the beauty of vocal music through personal practice, as well as clarify any problems or deficiencies in their own learning process, allowing them to improve and optimize their performance more effectively over time. Teachers can encourage students to express themselves more in their daily learning, and they can set up special sessions in the classroom to require students to sing relevant works in front of the podium one by one and to allow everyone to point out the advantages and disadvantages of singing to help students better know the connotation and content of the works. Teachers can also encourage students to express themselves more in their daily learning by providing them with opportunities to express themselves more in their daily learning. Significance: additionally, teachers should encourage pupils to engage in stage rehearsals and performances. By presenting their work to an audience and soliciting comments from that audience, students can clear the path for further progress of their work and establish favorable conditions for the continuing improvement of their overall quality [20, 21].

2.2.2. Optimizing the Course Teaching Plan.
The traditional single and backward curriculum teaching system seriously affects the learning interest of vocal music students and is not conducive to the cultivation of students' innovative ability. When carrying out vocal music teaching activities, teachers need to constantly innovate teaching methods in accordance with the requirements of the development of the times, establish an equal relationship between teachers and students, fundamentally stimulate students' subjective initiation, and enable students to actively participate in vocal music learning. Harmonious and pleasant classroom atmosphere: vocal music teachers should carry out teaching activities with students as the main body, encourage students to express their own views and attitude, and give students affirmations in a timely manner, so that students can gain a sense of achievement through continuous vocal music learning. For example, for classic vocal works, teachers can encourage students to actively carry out secondary creations, combining their own experiences and feelings to create more individual and innovative singing methods. Through their own understanding of the works, they integrate their own

emotions into the works and stimulate students' creativity. At the same time, teachers also need to continuously optimize and update teaching methods and content of teaching materials and use modern teaching methods to transform original boring theoretical knowledge into animation, music, or video, so that students can understand related concepts more quickly and accurately.

2.2.3. Forming a Professional and Reliable Teaching Team.
When it comes to effective vocal music education, a good vocal music teacher team is essential. First and foremost, vocal music teachers themselves must have a strong artistic concept for singing as well as correct singing abilities. Only professors who possess strong fundamental vocal abilities can give pupils with dependable educational materials. As a result, it is vital to increase the efficiency of the existing teaching staff while also raising the overall professional level of the teaching team.

Second, vocal music teachers are expected to have a particular level of proficiency in piano accompaniment, the capacity to sing with power and conviction onstage, and provide students with professional demonstrations in order to pique their interest and motivate them to continue their studies. Students can improve their own expressive notions by listening to the teacher's singing, which will allow them to comprehend and grasp the works at a deeper level of understanding. In addition, teachers must constantly broaden and deepen their knowledge of various subjects in order to be effective. In part, because the teaching of vocal music is a systematic process, it is impossible to thoroughly develop students' musical talent and musical success if you merely grasp the knowledge system associated with the subject of vocal music. As a result, teachers must continue to absorb theoretical knowledge in literature, history, language, aesthetics, and other areas and integrate it into their vocal music teaching, as well as continue to pursue related knowledge such as pedagogy and psychology, in order to further enrich and optimize the teaching system and content [22–26].

3. Vocal Management System and Melody Extraction Algorithm

3.1. Working Principle Analysis

3.1.1. Model Embedding System.
A critical aspect in communicating the influence of virtual reality experience in the early stages is the model embedding system, and its geometric segmentation function can play a significant role in the fluency of scene interaction experience in virtual reality. Thus, the same model is separated into three levels of LOD accuracy standards, allowing the model accuracy to be switched between different lines of sight ranges with greater flexibility. Generally, the model accuracy of the foreground should be high, the model accuracy of the medium scene should be moderate or expressed by a simple model, and the model accuracy of the distant view should be low and can even be replaced by texture masks or images, according to the principle that should be followed by this switching

method. The term is used to describe a technique that allows virtual reality to be optimized in real-time based on diverse viewpoints of the scene.

3.1.2. Face Capture System.
Face capture refers to the process of determining the size, position, distance, and other attributes of facial features such as the iris, nose, mouth corners, and other similar features and then calculating their geometric feature quantities to form a feature vector that describes the face as a whole. The primary premise of its technology is to use local human body feature analysis and neural recognition algorithms in conjunction with each other. The primary goal is to compare, judge, and confirm a series of processes with all of the original parameters in the recognition database generated by geometric relationship multidata using the features of human facial activities.

3.1.3. Gesture Capture System.
The gesture capture system is based on the Oculus quest2 hand positioning and tracking technology, which captures the spatial coordinates of the joints of the human hand and transmits them to the virtual reality animation process in real-time. Oculus quest2 is a hand positioning and tracking system that was developed by Facebook. Each finger's bending position is recorded, and the data normalization method is used to create a single-byte data format for all fingers, thereby reducing redundant data. Ensure that the bones and muscles of the gesture are in a relaxed, natural state of movement. In addition, during the process of gesture capture, segmentation algorithms based on obvious aspects will be developed, with the majority of them focusing on skin color segmentation and hand shape segmentation, among other things [18, 19].

3.1.4. Camera System.
In virtual reality, the camera system is typically input from the first perspective, which is also a sort of active vision calibration and can effectively record the dynamic vision witnessed by the human eye. In contrast to the classic virtual camera, it does not require the usage of a calibration item of known size to function properly. It is possible to record dynamic sequence frame images in time by establishing the corresponding method of coordinate points and image points on the calibration object, but only to a certain extent by doing so. When there is no stabilization function, you must perform the appropriate program optimization work at the bottom of the virtual reality program in order to obtain a steady camera function.

3.2. Melody Extraction System.
Figure 1 depicts the general framework for the automatic extraction of the main melody of a vocal performance. The entire algorithm is divided into four sections. The first section is used for audio preprocessing, note segmentation, and voiced segment detection. The second section is used for voiced segment detection. In the first step, the incoming music signal is normalized, decibels are added, and the time field is left

blank. The wave structure of the target audio clip is identified using an audio clip detection algorithm. The second step is to identify a number of potential fundamental frequencies to use. First, the pitch saliency function of the voiced segment audio is calculated using the optimized comb filter, and then numerous candidate fundamental frequencies are retrieved for each frame using the optimal comb filter. The Viterbi algorithm is used in the third section to track the dominant fundamental frequency track in each voiced segment, which is then displayed in the fourth section. The identification of the main melody of the song is the subject of the fourth section. The trained fundamental frequency discrimination model is used to statistically identify whether each section of the dominant fundamental frequency track is a singing tune or not using only the fundamental frequency track. If this is the case, it is reserved. The link becomes the central focus of the song.

Preprocessing of audio signals comprises downsampling, normalization, framing, windowing, and time-frequency domain modification, among other things. In music, the harmonic components of the human voice over 4 kHz are usually responsible for the majority of the sound.

Because the ratio is so small, the original music signal is downsampled to 8 kHz, which can help reduce the amount of math required in the future processing. The audio signal is steady for only a brief period of time, and it is also framed and windowed in addition to other features. For the purposes of this work, a Hamming window is employed, and 320 samples are captured for each frame of the signal. In this study, the time-frequency transformation of the signal is accomplished by the use of the short-time Fourier transform. Framework of the melody extraction system is shown in Figure 1.

For example, for a vocal main melody (little star), we first perform data preprocessing and then extract the candidate main melodies and identify the main melody type, which in turn facilitates teachers and students to directly input the music into the management system and output the corresponding music tracks.

Singing and instrumental accompaniment are the components of music. They are all made up of notes that last a specific amount of time. Each note has frequency and harmonic properties that are reasonably steady. There is a significant disparity between the portions. A note segmentation technique, the metric distance (DIS) approach introduced in [8], is employed in this research to segment the notes. The DIS metric distance is denoted by

$$\text{DIS} = \frac{\left(\mu_1 - \mu_2\right)^T \left(\mu_1 - \mu_2\right)}{tr\left(\sigma_1\right) + tr\left(\sigma_2\right)}. \tag{1}$$

where μ_1 and μ_2 represent the mean vector of the two audio features before and after, and $tr\left(\sigma_1\right)$ and $tr\left(\sigma_2\right)$ represent the trace of the covariance matrix of the two audio features, respectively.

The characteristic parameters in this paper are short-term amplitude spectrum. By sliding the data window by frame, the DIS distance function $\text{DIS}(t)$ about the number of frames t is calculated as follows:
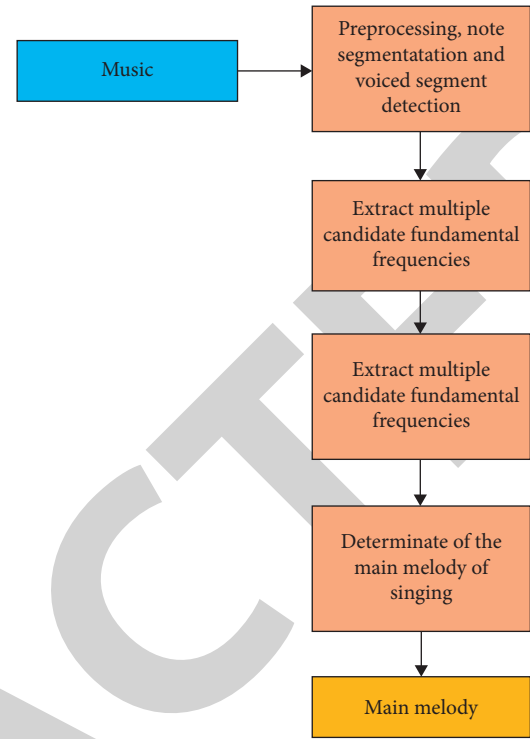

Figure 1: Framework of the melody extraction system.

$$\text{DIS} = \frac{\left(\mu_{t,1} - \mu_{t,2}\right)^T \left(\mu_{t,1} - \mu_{t,2}\right)}{tr\left(\sigma_{t,1}\right) + tr\left(\sigma_{t,2}\right)}, \tag{2}$$

where $\mu_{t,1}$ and $\mu_{t,2}$ represent the mean vector of the two audio features before and after the frame, and $tr\left(\sigma_{t,1}\right)$ and $tr\left(\sigma_{t,2}\right)$ respectively represent the trace of the covariance matrix of the two audio features before and after the frame.

Find all the maximum points in $\text{DIS}(t)$, set the threshold T_1 as the mean of $\text{DIS}(t)$, and delete the maximum points smaller than the threshold T_1. In addition, the duration of a quarter note in fast-paced music is about 0.5 s. Considering that the duration of an eighth note and a sixteenth note is 1/2 and 1/4 of that of a quarter note, the paragraph spacing is set to be no less than 100 ms; otherwise, the corresponding maximum point is removed, so that the remaining maximum point is the note split point [17, 20].

A voiced signal whose fundamental frequency is f_0 and its frequency domain expression is

$$Y(f) = \sum_{k=1}^{K} a_k \alpha\left(f - kf_0\right), \tag{3}$$

where a_k is the coefficient of the kth harmonic.

The logarithmic frequency domain can be expressed as

$$Y(q) = \sum_{k=1}^{K} a_k \alpha\left(q - \log k - \log f_0\right), \tag{4}$$

where $q = \log f$. In the logarithmic frequency domain, the spacing between the harmonics and f_0 is irrelevant, so when it is convolved with a filter, its impulse response is

$$h(q) = \sum_{k=1}^{K} \alpha(q - \log k). \tag{5}$$

The convolution result $Y(q) * h(q)$ will generate a peak at the position of $q_0 = \log f_0$, and the voiced fundamental frequency can be determined from the position of the peak. (5) is an ideal filter composed of many à functions. In practice, due to the windowing analysis, the width of the harmonic peak will be broadened, so the actual filter used is shown in

$$h(q) = \frac{1}{\beta - \cos(2\pi \exp(q))}. \tag{6}$$

In order to obtain the comb filter illustrated in Figure 2, select the required filter parameters. Convolute the logarithmic domain spectrum of a frame signal with the comb filter in order to obtain the fundamental frequency saliency function of the frame signal:

$$S(q) = Y(q) * h(-q). \tag{7}$$

After extracting the multiple candidate fundamental frequencies, the dominant fundamental frequency trace is tracked within a segment (within each note) using the Viterbi algorithm. In this paper, the Viterbi algorithm adopts the pitch likelihood and pitch transition probability, and the pitch likelihood is defined as

$$p(f_m | X_t) = \frac{s_{t,m}}{\sum_n s_{t,m}}. \tag{8}$$

where X_t is the amplitude spectrum of the t-th frame, $s_{t,m}$ is the pitch saliency value of the m-th candidate fundamental frequency f of the t-frame, $\sum_n s_{t,m}$ is the pitch saliency value of all the candidate fundamental frequencies of the t-th frame, sum of degree values.

The pitch transition probability in the text is obtained from the statistics of the annotated music library. Define the pitch change rate Δf of adjacent frames as

$$\Delta f = \frac{f_t - f_{t-1}}{f_{t-1}}. \tag{9}$$

After extracting the dominant fundamental frequency track from each note segment, determine whether it belongs to the singing voice or the accompaniment (voiced segment). Keep the dominant fundamental frequency track if it pertains to the singing voice; delete it if it pertains to the accompaniment. For this purpose, this section includes an addition of a fundamental frequency discriminant model.

$$S = \sum_i \lfloor y_{ij} \times MSE(y_i) \rfloor. \tag{10}$$

The main melody discrimination implementation steps:

(1) Construct a comb filter with a frequency range of 0~4 kHz from the dominant fundamental frequency F_0, as shown in

$$h(f) = \sum_{k=1}^{K} \alpha(f - kF_0) * b(f). \tag{11}$$
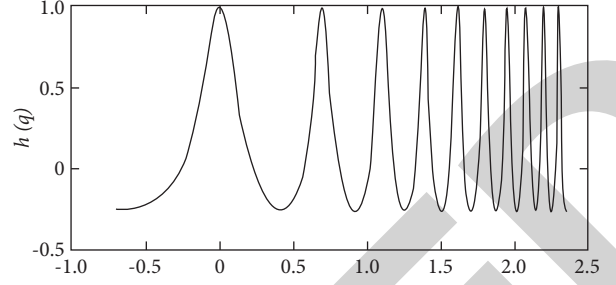

Figure 2: Parametric statistics of Juniper.

where K is the number of harmonics in the range of 0~4 kHz, and $b(f)$ is the rectangle waveform of the comb filter.

(2) Filter the signal amplitude spectrum with a comb filter to obtain the harmonic spectrum corresponding to F_0, and extract its corresponding MFCC parameters.

(3) Send the Mel frequency cepstral coefficients to the neural network for identification and judge whether the dominant fundamental frequency F_0 is the singing fundamental frequency.

(4) When analyzing each voiced segment, count how many frames are devoted to the fundamental frequency of the singing voice, and if this number is greater than half the total number of frames devoted to the voiced segment, it can be concluded that the dominant fundamental frequency track in that voiced segment is responsible for the main melody of the singing voice [24].

4. Results

The music data is taken from the MIR-1K data set, which contains 1,000 pieces of music with a sampling rate of 16 kHz, the ability to distinguish the singing voice from the accompaniment, and the ability to label the fundamental frequency of the singing voice with a time interval of 10 ms. This experiment utilizes the MIR-1K data set, with 500 pieces of music chosen at random as the training set and the remaining 500 pieces of music being utilized as the primary melody to extract the test set from the remaining 500 pieces of music. In this paper, 75 percent is divided into the training set and the rest is the test set.

This section conducts experiments with 500 pieces of music from the test set as the test data in order to verify the accuracy of the automatic extraction algorithm of the vocal main melody. The signal-to-interference ratio (SIR) is set to 0 dB and 5 dB, respectively, in order to verify the accuracy of the automatic extraction algorithm of the vocal main melody.

There are two basic tasks in vocal theme extraction: one is to judge whether the melody really exists, and the other is to accurately estimate the pitch of the theme. Focusing on the basic target task, the performance of the algorithm is also evaluated by five performance indicators used in [12], melody localization recall rate (VR), melody localization
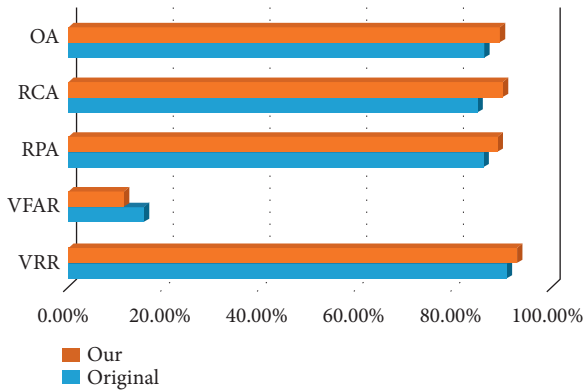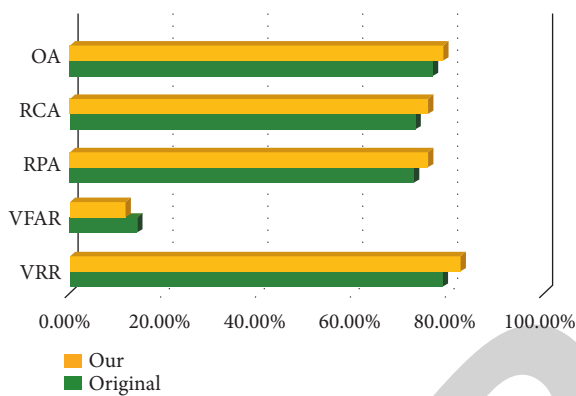
Figure 3: Result with SIR = 5 db.
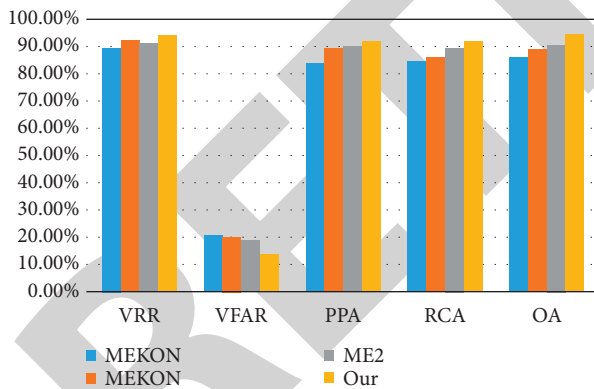

Figure 4: Result with SIR = 0 db.

Figure 5: Different model comparison results with SIR = 5 db.

false alarm rate (VFAR), original pitch accuracy (RPA), original color accuracy (RCA), and overall accuracy (OA).

The test data set consists of a random sample of 500 pieces of music, and the following tests are conducted with a signal-to-interference ratio of 0 dB and 5 dB for the first and second experiments, respectively. Comparing the original algorithm described in Reference 5 to the algorithm described in this paper yields the specific results depicted in Figures 3 and 4.

Additionally, we compare the algorithm in this paper to the literature's MEKON, ME, and ME2 algorithms under
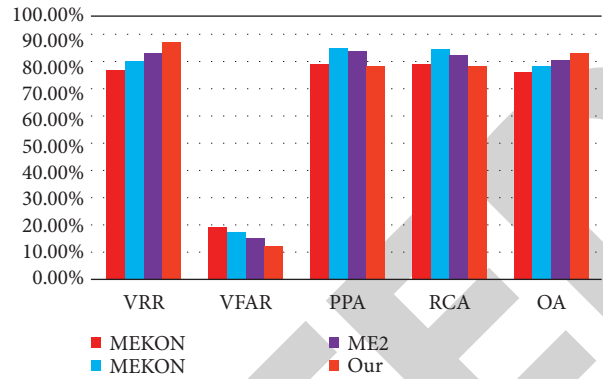

Figure 6: Different model comparison results with SIR = 0 db.

various SIRs, as illustrated in Figures 5 and 6. As illustrated in Figures 3–6, the algorithm used in this paper is effective at extracting melodies.

In terms of time-consuming, our method takes the shortest time regardless of SIR of 0 dm or 5 dm.

Further, we did the Fridman test and obtained $P < 0.05$, indicating that our method significantly outperforms the comparison algorithm.

5. Conclusion

This study presents a management system for vocal music instruction through the use of virtual reality technology. Additionally, this study presents an improved algorithm for automatically extracting the main melody from a vocal performance. Taking into account the spectral characteristics of vocal music signals, the pitch saliency function calculation technique has been modified to reduce the computational complexity and time required to extract the primary melody from the vocal music signal. The model presented in this study has the potential to significantly improve the recognition accuracy of the main melody model, the rate of false alarms during melody localization, and the overall accuracy of vocal main melody extraction. Additionally, this research incorporates the extraction technique into the management system, which makes it easier for instructors to use in the classroom [27–29].

Data Availability

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

Conflicts of Interest

The author declares that he has no conflicts of interest.

References

[1] J. Salamon, E. Gomez, and D. P. W. Ellis, "Melody extraction from polyphonic music signals: approaches, applications, and challenges[J]," *IEEE Signal Processing Magazine*, vol. 31, no. 2, pp. 118–134, 2014.

[2] X. F. Zheng, *Research on the Whole Teaching of Vocal Music Course in University Music Performance Major Based on*

Multimedia Technology[J], p. 2022, Scientific Programming, USA, 2022.

[3] W. H. Tsai and H. M. Wang, "On the extraction of vocal-related information to facilitate the management of popular music collections[C]," in *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 197–206, UK, 2005.

[4] D. Gebre-Egziabher, G. H. Elkaim, J. David Powell, and B. W. Parkinson, "Calibration of strapdown magnetometers in magnetic field domain," *Journal of Aerospace Engineering*, vol. 19, no. 2, pp. 87–102, 2006.

[5] W. M. Frix, G. G. Karady, and B. A. Venetz, "Comparison of calibration systems for magnetic field measurement equipment," *IEEE Transactions on Power Delivery*, vol. 9, no. 1, pp. 100–108, 1994.

[6] M. Euston, P. Coote, R. Mahony, J. Kim, and T. Hamel, "A complementary filter for attitude estimation of a fixed-wing UAV," in *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 340–345, IEEE, September 2008.

[7] G. F. Welch, D. M. Howard, and C. Rush, "Real-time visual feedback in the development of vocal pitch accuracy in singing[J]," *Psychology of Music*, vol. 17, no. 2, pp. 146–157, 1989.

[8] E. Cano, G. Schuller, and C. Dittmar, "Pitch-informed solo and accompaniment separation towards its use in music education applications[J]," *EURASIP Journal on Applied Signal Processing*, vol. 2014, no. 1, pp. 1–19, 2014.

[9] P. Knees and M. Schedl, "Introduction to music similarity and retrieval," *Similarity and Retrieval*, Springer, Berlin, Heidelberg, pp. 1–30, 2016.

[10] K. Hoashi, K. Matsumoto, and N. Inoue, "Personalization of user profiles for content-based music retrieval based on relevance feedback[C]," in *Proceedings of the Eleventh ACM International Conference on Multimedia*, pp. 110–119, Germany, 2003.

[11] R. Mahony, T. Hamel, and J. M. Pflimlin, "Complementary filter design on the special orthogonal group SO(3)," in *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 1477–1484, IEEE, Seville, Spain, December 2005.

[12] L. Lu, M. Wang, and H. J. Zhang, "Repeating pattern discovery and structure analysis from acoustic music data[C]," in *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 275–282, France, 2004.

[13] R. C. Luo, M. H. Lin, and R. S. Scherp, "Dynamic multi-sensor data fusion system for intelligent robots," *IEEE Journal on Robotics and Automation*, vol. 4, no. 4, pp. 386–396, 2007.

[14] P. Rao, *Audio, Image and Biomedical Signal Processing Using Neural Networks*, pp. 169–189, Springer, Berlin, Heidelberg, 2008.

[15] L. Lu and H. J. Zhang, "Automated extraction of music snippets," in *Proceedings of the Eleventh ACM International Conference on Multimedia*, pp. 140–147, USA, 2003.

[16] O. Cornelis, M. Lesaffre, D. Moelants, and M. Leman, "Access to ethnic music: advances and perspectives in content-based music information retrieval[J]," *Signal Processing*, vol. 90, no. 4, pp. 1008–1031, 2010.

[17] M. J. Bergee, R. L. Tast, and B. Wheeler, "AQ analysis of judgments of centrality in music education courses of study [J]," *Bulletin of the Council for Research in Music Education*, vol. 2, no. 225, pp. 7–21, 2020.

[18] S. Wu, "Research on the development of pattern recognition and applications on digital music recognition and creation," in *Proceedings of the 2016 2nd International Conference on Social Science and Technology Education (ICSSTE 2016)*, pp. 646–651, Atlantis Press, Russia, 2016.

[19] T. B. Scott, *The Construction of a Holistic, Criterion-Referenced Sight-Singing Test for High School Sopranos Based on the Voluntary National Standards for Music education[M]*, University of Illinois at Urbana-Champaign, USA, 1996.

[20] N. Kroher and J. M. Díaz-Báñez, "Modelling melodic variation and extracting melodic templates from flamenco singing performances[J]," *Journal of Mathematics and Music*, vol. 13, no. 2, pp. 150–170, 2019.

[21] M. Lesaffre, M. Leman, B. De Baets, and J. pierre Martens, "Methodological considerations concerning manual annotation of musical audio in function of algorithm development," in *Proceedings of the 5th International conference on Music Information Retrieval*, pp. 64–71, Barcelona, Spain, 2004.

[22] K. Burwell, "On musicians and singers. An investigation of different approaches taken by vocal and instrumental teachers in higher education[J]," *Music Education Research*, vol. 8, no. 3, pp. 331–347, 2006.

[23] G. Kamberova, R. Mandelbaum, J. S. De Almeida, M. Mintz, and R. Bajcsy, "Decision-theoretic approach to robust fusion of location data," *Journal of the Franklin Institute*, vol. 336, no. 2, pp. 269–284, 1999.

[24] P. Li, X. Wang, N. Li, X. Wang, X. Lizhi, and Z. Mingquan, "A novel MIR framework and application with automatic voice processing, database construction and fuzzy matching," *IEEE* in *Proceedings of the 2010 the 2nd International Conference on Computer and Automation Engineering (ICCAE)*, vol. 1, pp. 20–24, Singapore, 2010.

[25] M. Y. Liao and P. S. Campbell, "Teaching children's songs: a Taiwan–US comparison of approaches by kindergarten teachers[J]," *Music Education Research*, vol. 18, no. 1, pp. 20–38, 2016.

[26] T. Zhang and C. C. J. Kuo, "Heuristic approach for generic audio data segmentation and annotation[C]," in *Proceedings of the seventh ACM international conference on Multimedia*, pp. 67–76, Italy, 1999.

[27] J. Salamon and J. Urbano, "Current challenges in the evaluation of predominant melody extraction algorithms[C]," in *Proceedings of the ISMIR*, vol. 12, pp. 289–294, Australia, 2012.

[28] L. Xu, Y. Chen, and P. Cui, "Improvement of DS evidential theory in multisensor data fusion system," in *Proceedings of the Fifth World Congress on Intelligent Control and Automation*, pp. 3124–3128, IEEE, Hangzhou, China, June 2004.

[29] H. Qi and J. B. Moore, "Direct Kalman filtering approach for GPS/INS integration," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 678–693, 2002.