International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-19

https://doi.org/10.70917/ijcisim-2026-1821
Article

A study on the comprehensive development of vocal education under the dual role of voice training and mental quality cultivation

Shifang Yang[1,*]

[1] School of Preschool Education, Chongqing Youth Vocational & Technical College, Chongqing, 400700, China
* Correspondence author: 13677645698@163.com

Abstract: The development of modern vocal education presents a diversified trend, in which the application and development of artificial intelligence technology has a positive impact on voice training and mental quality training. In this paper, we first use the average amplitude difference function method to extract the fundamental tone of the music signal, and find that there is a misdetection of the fundamental tone in the process, which is smoothed by using the median smoothing algorithm and the linear smoothing algorithm mixed method, aiming at eliminating the interfering information in the music signal. On this basis, by improving the vocal feature matching of the DTW algorithm, the intelligent diagnosis of voice training in vocal music education is realized, with a view to improving the quality of students' voice training. According to the actual performance of students in the process of voice training, it is proposed to use the Bayesian Knowledge Tracking-based Music Resource Recommendation Model (BKT-ER) to help students master the knowledge points of vocal training, which prompts students to overcome the psychological pressure and bad emotions in playing, and helps to establish a good psychological quality. Finally, from the voice training and psychological quality training program, the integration of the two application program is developed, and an example analysis of the program is carried out. Under the dual role of voice training and psychological quality cultivation in vocal education, students' basic skills, emotional regulation ability, and psychological cognitive ability are effectively improved, and the mean value of its dimensions is greater than 3.5, which fully verifies the effectiveness of the fusion program in this paper.

Keywords: average magnitude difference function method; improved DTW algorithm; BKT-ER; vocal education; voice training; psychological quality training

1. Introduction

Vocal music is a comprehensive art, and he has high requirements for students' musical skills and practical ability. In vocal music education, it is not only necessary to master the basic musical skills, but also to cultivate a good voice, cultural cultivation, psychological quality and so on, in order to perfectly show and express the music, so that the rich cultural connotation and spiritual essence of music can be carried forward [1-3]. Among them, voice training and psychological quality training are important factors in promoting the comprehensive development of students.

Good vocal art skills require vocal singers to have more excellent vocal conditions, which is conducive to conveying the emotions of vocal works, promoting the standardized development of vocal art, and is an essential quality in the comprehensive development of vocal music [4-6]. A good voice needs to be formed through continuous practice and strict professional training, and the singing skills should be gradually mastered in continuous practice [7-8]. However, in the comprehensive development of vocal education, the cultivation of psychological quality is the most neglected important issue. Psychological research shows that any human activity is carried out under

psychological regulation, and psychological activity is directly related to human practical ability [9]. Although singing is a combination of multiple actions under the joint participation of human breathing, vocal resonance, biting and spitting, hearing, etc., when singing, the good or bad psychological state or psychological quality of a person will play a very important role in the training and learning of their singing, and in the advantages and disadvantages of the teacher's teaching results [10-13]. Therefore, in vocal music teaching should pay attention to the cultivation of psychological quality of students, according to the intrinsic characteristics of vocal art, actively explore more scientific and effective teaching methods, and constantly update the concept of education in order to improve the quality of teaching [14-15].

To address the importance of psychological quality training in vocal music education, literature [16] analyzed the existing problems of insufficient demonstration and low motivation in vocal music classroom teaching through a questionnaire survey, and examined the effectiveness of the application of strategies such as emotional cues and infections on the learning effect based on the theory of positive psychology, and emphasized that integrating the emotional experience into the skills training is a key path to improve the effectiveness of teaching. Literature [17] analyzed the abstractness and complexity of vocal music teaching and pointed out that it is difficult to achieve good results only by skills training, and emphasized that psychological factors have a significant impact on the effectiveness of teaching and learning, and that it is necessary to continue to cultivate the ability of psychological regulation in order to promote the overall improvement of the quality of vocal music education. Literature [18] introduced a comparative approach to the reform of the functionalist principle of the path and combined with the support vector machine algorithm to analyze the effectiveness of its intervention on the emotional relief of patients with anxiety, pointing out that the program can effectively improve the psychological problems such as anxiety, depression, and fear in order to enhance the level of mental health. Based on social cognition and flow theory, [19] examined the impact of vocal music education on students' psychological performance in Shandong University of the Arts, analyzed the psychological factors such as motivation and self-efficacy, pointed out that the integration of Chinese culture can enhance psychological resilience and identity, and emphasized that the construction of a supportive learning environment has a key role to play in improving the psychological quality and comprehensive ability. By analyzing the decisive role of psychological factors in vocal music learning, [20] pointed out that the common bad state of mind would restrict the singing performance, and emphasized that strengthening the skills practice and stage practice to improve the artistic quality and psychological control ability is the key path to cultivate a good singing psychology. Literature [21] analyzes the intrinsic connection between educational psychology and Chinese folk vocal music education and examines the path of optimizing teaching design with psychological laws, pointing out that teachers can accurately grasp the physical and mental changes of students with the help of a psychological perspective to provide targeted guidance, and emphasizing that integrating psychological education into vocal music skill training is a key strategy to improve artistic literacy and values. Literature [22] analyzed the relationship between entrepreneurial self-efficacy and entrepreneurial intention in music education in colleges and universities based on questionnaires and path models, pointed out that music education can enhance entrepreneurial self-efficacy, but it fails to drive entrepreneurial intention and gender plays a moderating role, and emphasized that the cultivation of mental qualities and comprehensive competitiveness is the key to improve the innovation and entrepreneurship education model. Literature [23] used interviews and data analysis to examine the impact of singing psychological quality on stage performance, pointed out that only focusing on skill training but neglecting psychological adjustment will lead to clinical tension and play failure, and emphasized that integrating psychological quality training system into vocal music courses is a key path to improve singing quality and teaching effectiveness. Literature [24] discusses the multiple factors affecting the psychological construction of singing, such as personal psychological quality and external environment, and points out that a good singing state is the foundation of successful singing, and emphasizes that vocal educators must understand the causes of psychological barriers in order to optimize the teaching methods to enhance students' self-confidence and artistic expression. Literature [25] examined the effects of music education and sports psychology on college students' anxiety through an eight-week intervention experiment. The analysis pointed out that the combination of sports prescription and group counseling could significantly increase the levels of 5-HT and BNDF to alleviate anxiety, and emphasized that 37% of the students had symptoms of anxiety, which should be highly valued by colleges and universities. Literature [26] examined the role of psychological regulation in vocal music teaching on the cultivation of learning motivation and willpower, pointed out that it can enhance interest and teaching effectiveness, and emphasized that the integration of emotional experience into psychological training is a key path to maintain good singing status and avoid bad emotions. The above study reveals the key role of psychological quality development in vocal music education, and that

students' psychological resilience, self-efficacy, artistic performance, and mental health can be effectively enhanced through the multiple paths of positive psychology, functionalist interventions, social cognitive theories, the cultivation of psychological regulation, and co-counseling.

This paper refers to the current mainstream several gene extraction algorithms, and finally chooses to use the average amplitude difference function method to carry out the base note extraction work of the music signal, and finds that there is interference information in this process, for this reason, adopts the median smoothing algorithm and linear smoothing algorithm mixed method to smooth the processing, in order to ensure the usability of the research data. After completing the data preprocessing work, the improved DTW algorithm was used to match the musical features, aiming to realize the intelligent diagnosis of voice training in vocal music education. Subsequently, the Bayesian Knowledge Tracking-based Music Resource Recommendation (BKT-ER) model is proposed to help students' mental quality training by combining the students' knowledge mastery in the theory during voice training. Finally, taking voice training and psychological quality training in vocal education as the entry point, we formulate a program for the integration of the two, aiming to improve the comprehensive development level and quality of vocal education in colleges and universities.

2. Voice Training and Mental Quality Cultivation in Vocal Music Education

2.1. Voice training in vocal education

With the development of network and digital technology, in China's vocal education, the introduction of digital music teaching on the basis of the previous general vocal teaching means, and as a supplementary teaching method has become an inevitable trend. This auxiliary teaching means can use digital technology to simulate the sound of strings, wind, percussion and so on, but also can fully display the rock, jazz, tango and other rhythms, which brings opportunities for voice training in vocal education.

2.1.1. Base note extraction

Sound has the physical quality of acoustic characteristics, and its acoustic characteristics are timbre, pitch, duration and intensity, referred to as the four elements of sound. Pitch is an important characteristic parameter for evaluating the features in this paper. The sound pitch feature is converted from the gene profile feature. There are three main methods summarized for base pitch extraction, namely, the time domain method, the frequency domain method, and the combined time and frequency domain method. The time-domain method has several main methods, i.e., short-time autocorrelation function method (STACF), zero-crossing rate method (ZCR), parallel processing technique (PPROC), average magnitude difference function method (AMDF), harmonic superposition method (SHS), and so on. The frequency domain methods can also be categorized into several major methods, i.e., Simple Inverse Filter Tracking (SIFT) method, Frequency Extreme Point Detection (FEPD) method, FIR filter, Wavelet Function Method (WFM) and so on. In addition to these methods mentioned above, there are also instantaneous frequency methods, Comb transform methods, inverse spectral methods and so on. Among them, the short-time autocorrelation function method, the average magnitude difference function method, and the inverse spectrum method are commonly used algorithms for base note feature extraction, while the average magnitude difference function method (AMDF) is mainly used in this paper.

The short-time average amplitude difference function of the music signal $\{s(n)\}$ is defined as:

$$F_n(\tau) = \frac{1}{R}\sum_{n=0}^{N-1}\left|s(n+m)w_1(m) - s(n+m+\tau)w_2(m+\tau)\right| \tag{1}$$

In the above function, $w(m)$ refers to a window function, N characterizes the actual length of a frame of music signal, and we can think of the period of the fundamental as P, mainly because the turbulent segments of the music signal are all periodic, so we can think of the turbulent segments, i.e., $F_n(\tau)$, at the points $\tau = +P, +2P, +3P\ cdots$ are the locations of the troughs, and the fundamental period is the interval between these troughs.

In practice, the window functions $w_1(m)$ and $w_2(m)$ we take are preferably both rectangular windows to make the short-time average amplitude difference function $F_n(\tau)$ easier to express and compute. In this case, the short-time average amplitude difference function is expressed as:

$$F_n(\tau)=\frac{1}{R}\sum_{n=0}^{N-1}\left|s(n)-s(n+\tau)\right| \ \tau=0,1,\cdots N-1 \tag{2}$$

From this function above, we can see that the function $F_n(\tau)$ carries out relatively simple calculations, only addition, subtraction and absolute value operations, which can be realized when carrying out the design of software and hardware, so the short-time average magnitude difference function method is very widely used when carrying out the detection of fundamental tones, for example, the 10th-order linear prediction vocoder we commonly use, the LPC-10 vocoder, can be used to carry out a very good extraction of the The fundamental tone period is well extracted, and this device is designed based on the AMDF method. It is proved through experiments that when the environment is silent or the noise is relatively small, the AMDF method is more effective in the extraction of the fundamental cycle, but when the music environment is worse, the noise is more, and the signal-to-noise ratio is small, the detection effect is not so satisfactory, and it is necessary to take other methods to carry out.

Figure 1 is a randomly selected frame of the vocal signal, this frame of the vocal signal to carry out the calculation of the average amplitude difference function, the average amplitude difference function is shown in Figure 2, as can be seen from the figure, the average amplitude difference function at the fundamental period shows the amount of the troughs, and the fundamental period is the value of the distance between these troughs.

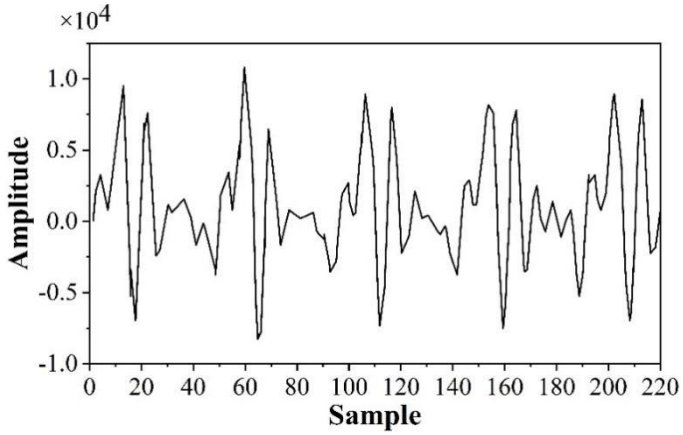


Figure 1. Original signal

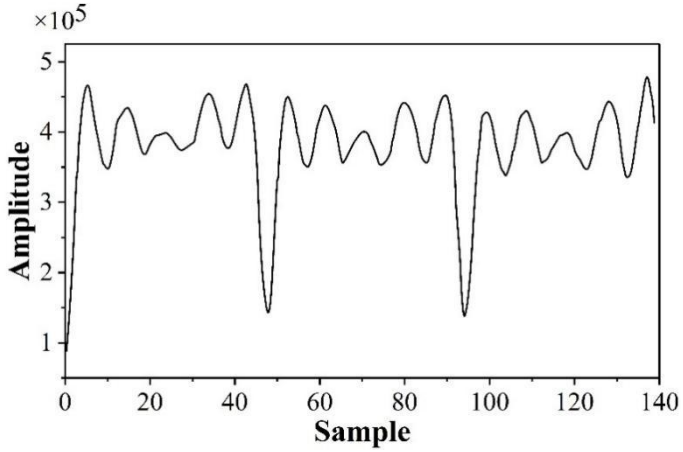


Figure 2. Average amplitude difference function

When noise with a signal-to-noise ratio of 2 dB is added to this frame of music signal, the added noise signal is shown in Fig. 3. Similarly, the average amplitude difference function is also calculated for this frame of music signal with added noise, and the image of the average amplitude difference function is shown in Fig. 4, from which it can be seen that the average amplitude difference function increases a lot of harmonic components, and in these harmonic components, the peak of the fundamental cannot be found at all, and the fundamental period cannot be judged at all.

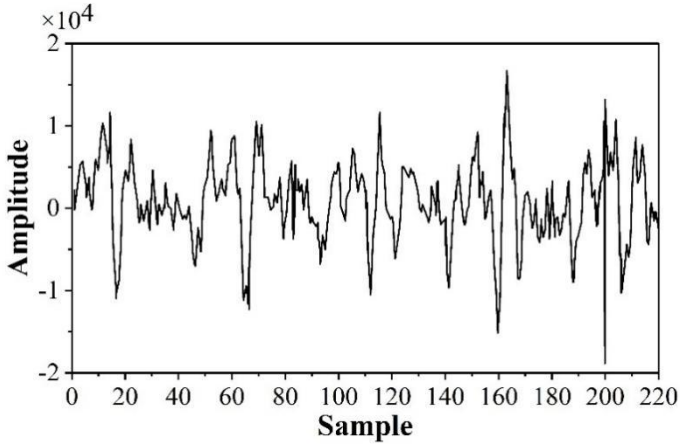


Figure 3. Noisy signal

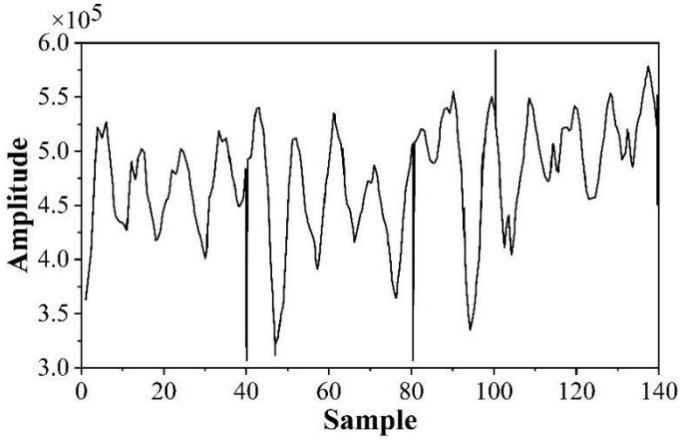


Figure 4. Average amplitude difference function graph

2.1.2. Base note smoothing

Although there are so many methods for detecting fundamental cycles, no matter what method is used, there can be false detection of the fundamental, making one or several fundamental cycles in the trajectory of the fundamental cycle and the previous estimate of the value of the gap is very large, generally speaking, there will be twice or 1/2 the deviation from the normal value, this deviation produces a point of deviation that we generally become the fundamental trajectory of the " wild spots". The existence of these wild points is unnecessary, must be eliminated, this time with the help of a variety of smoothing algorithms, and in these algorithms, the most widely used is the median smoothing algorithm and linear smoothing algorithm, this paper adopts the median smoothing algorithm and linear smoothing algorithm mixed for processing.

(1) Median smoothing algorithm

The basic principle of the median algorithm can be summarized as follows: assuming that $x(n)$ is a random input signal, and $y(n)$ is the output signal of the input signal after passing through the median filter, the method adopted is usually to add a sliding window, the center of the window will be moved to the position of n the median of the window's input sample point as well as the left and right positions of the n point to take the value of L each value is the so-called output value. the so-called output value. The centers of these smoothing points can form a set of signal points with a length of $2L+1$ samples, which then form a sequence in order of size, and the point in the middle of the resulting sequence is the point we need, which is the so-called output of the smoother. The value of L is typically 1 or 2, which means that the median smoothing window contains typically 3 or 5 points during the operation. The advantage of median smoothing is mainly reflected in its signal integrity, based on the minimization of the wild points, the base tone cycle trajectory of the two smooth segments close to each other will not appear step, making the signal distortion.

(2) Linear Smoothing Algorithm

Linear smoothing processing uses a sliding window for linear filtering processing, i.e.:

$$y(n)=\sum_{m=-L}^{L}x(n-m)*w(m) \tag{3}$$

where $\{w(m),m=-L,-L+1,\cdots,0,1,2,\cdots,L\}$ is a smoothed window of $(2L+1)$ points that satisfies: $\sum_{m=-L}^{L}w(m)=1$. Linear smoothing not only minimizes the value of "wild points" in the signal, but also modifies the value of the sample points to some extent. In this case, the length of the window is increased, then the smoothing effect will become more obvious, but this will also make the step phenomenon between the two neighboring smoothing segments will be aggravated, the signal will be more blurred.

Therefore, in order to make the smoothing effect more obvious, the general method is to connect the two medians, the combination of the two medians is smoother, which is the use of median smoothing and linear smoothing algorithms, which can make the smoothing effect more obvious. After these processes, the smoothed fundamental trajectory will become closer, so it is necessary to carry out a second smoothing to carry out the process, after the first balancing, the fundamental trajectory will be smoothed even further.

2.1.3. Feature Matching

On the basis of completing the work of base note extraction and base note smoothing in vocal education, the improved DTW algorithm is used for vocal feature matching, aiming at realizing the comparison between the singers' voices and the corresponding data in the standard library, giving feedback to the students on the results, finding out the errors, and facilitating the students to grasp the situation of their own voice training. The specific process is as follows:

(1) Traditional DTW algorithm

Dynamic time regularization algorithm (DTW) is based on the idea of dynamic planning, creatively combining the calculation of distance measure and dynamic time regularization method effectively, which is a nonlinear regularization technique, solving the template matching problem of different vocal lengths, and can find the optimal correspondence between two given sequences, and the two time sequences are aligned as shown in Figure 5. In the fields of data discovery and information retrieval, DTW has been successfully applied to automatically process time-dependent time-varying or speed-varying data.

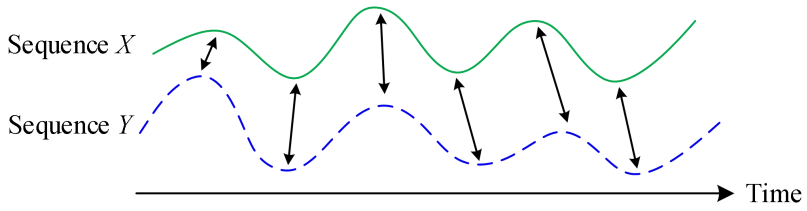


Figure 5. Schematic diagram of alignment of two time series

Define the audio input as the test template $P=\{chro_p(i),1\le i\le M\}$, M is the number of performance audio features. The score audio is the reference template $S=\{chro_s(j),1\le j\le N\}$, N is the number of score audio features. The time axis i of the test vector is mapped nonlinearly onto the time axis j of the reference template.

Compare the similarity between them by comparing the distances; if the distance between them is smaller, then the similarity is higher and vice versa. In order to compare the similarity between them, the distance between them can be compared, the smaller the distance the higher the similarity. Calculate the Euclidean distance between the performance audio P eigenvalue vector $chro_p(i),i=1\cdots N$ and the score audio S eigenvalue vector $chro_s(j),j=1\cdots M$, which has the advantages of being simple, fast, satisfying the triangular inequality, and supporting various indexing techniques. The similarity matrix (SM) is obtained as shown in equation (4):

The recursive idea is used to compute the cumulative matrix D, as shown in equation (4). Where $D(1,1)=SM(1,1)$, the dynamic planning path itself is obtained from $D(i,j)$. Namely:

$$D(i,j)=SM(i,j)+\min(A,B,C)=SM(i,j)+\min\begin{Bmatrix} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{Bmatrix} \tag{4}$$

Termination operation: $D(P,S)=D(M,N)$

(2) Improved DTW algorithm

The time accuracy of the traditional DTW algorithm is determined by the length of the subframe, shortening the length of the frame can get high accuracy, but the system running time grows. Moreover, regardless of the length of the sub-frame, it is impossible to accurately locate where the deviation note position is located. In order to make up for the shortcomings of traditional algorithms, this paper proposes a DTW algorithm combined with endpoint detection.

Note onset is a very important timing feature of music, and the timing information of music is crucial for analysis, labeling, etc. Accurate onset information can help locate important timing information such as music tempo, velocity, etc., and moreover, it can be used to segment the music into units of notes for processing, replacing the traditional continuous processing in terms of frames to improve efficiency.

The current mainstream algorithm for endpoint detection is to look for abrupt transition regions in the signal, i.e., regions where there is a sudden increase in the energy or energy spectrum of the signal. However, in chordal music where several notes occur at the same time, there may be a masking effect that does not detect the energy surge, making the detection method ineffective. In order to avoid the masking effect, this paper decomposes the frequency band into several non-overlapping subbands and extracts the signal transients within a certain frequency range.

The audio signal is short-time smooth, so it is analyzed before the processing of windowing and framing, after which the Fourier transform of a certain frame is calculated, i.e., the short-time Fourier transform. Through the short-time Fourier transform, the spectrum of the audio signal is obtained $X=\left(X\left(t,k\right)\right)_{t,k}$, $k=1,2\cdots,K$, $t=1,2\cdots,T$ where K is the number of samples in each frame, and T is the number of frames of the signal, and the frame length is chosen to be 0.023s. as the frame length. The frequency band is divided into $[0\ 500]$, $[500\ 1250]$, $[1250\ 3125]$, $[3125\ 7812.5]$, $[7812.5\ f_s/2]$, f_s is the sampling frequency, the spectral amplitude of each sub-band $|X|$ for the logarithmic operation, obeying the $Y=\log\left(1+C\cdot|X|\right)$, with constant $C=1000$. This not only adapts to the logarithmic relationship of the note frequencies, but also adjusts the dynamic range of the signal and enhances the clarity of the weaker transients in the high-frequency region.

To obtain the endpoint intensity profile, the discrete derivative of the compression spectrum Y is calculated as shown in Eq. (5):

$$\Delta(t)=\sum_{k=1}^{K}\left|Y(t+1,k)-Y(t,k)\right|_{\geq 0},t\in Z,\left|x\right|_{\geq 0}=\begin{cases} x & x\geq 0 \\ 0 & x<0 \end{cases} \tag{5}$$

To address the above problems, this paper adopts the constant Q transform (CQT), which has the same exponential distribution law of spectral frequency and scale frequency, in the time-frequency transform stage when processing music audio signals.CQT is a transform with a constant frequency bandwidth ratio, and the frequency resolution can be transformed along with the transformation of the frequency, which is more suitable for the characteristics of the frequency domain of music signals, and is able to better discriminate between bass-region scales, and possesses many of the basic qualities of the DFT, which is defined as shown in Equation (6):

$$f_k=2^{k/\beta}f_{\min} \tag{6}$$

where f_k is the scale frequency, representing the k th frequency component of the music signal within the transformed spectrum in Hz, $f_{\min}$ is the lower frequency limit of the processed music signal, and the lower limit in this paper is chosen as $f_{\min}=73.42\text{Hz}$. β is the number of frequency spectral lines contained in an octave, which determines the sampling rate of the algorithm.

Define the frequency bandwidth ratio as Q, a constant determined by β that satisfies Equation (7).CQT is named precisely because it keeps Q constant. Namely:

$$Q = \frac{f}{\delta_f} = \frac{f_k}{f_{k+1} - f_k} = \frac{1}{2^{1/\beta} - 1} \tag{7}$$

where δ_f is the frequency resolution and represents the bandwidth at frequency f . The $\beta = 36$ represents 3 frequency spectral lines within each semitone. Frequency resolution is defined as the ratio of the sample rate to the window length, from which equation (8) can be derived:

$$N_k = \frac{f_s}{\delta_{f_k}} = Q\frac{f_s}{f_k} \tag{8}$$

where N_k is related to the k value of the frequency domain subscript of the CQ spectrum, which is the window length that varies with frequency, and f_s is the sampling frequency.

Finally, the k th component of the CQT can be obtained by modeling the calculation of the corresponding frequency component in the DFT, as shown in Eq. (9):

$$X^{cq}(k) = \frac{1}{N_k}\sum_{n=0}^{N_k-1} x(n) w_{N_k}(n) e^{-j\frac{2\pi Q}{N_k}n} \tag{9}$$

where $x(n)$ is the time-domain signal and $w_{N_k}(n)$ is a window function of length N_k .

CQT adopts an isoperimetric sequence with a constant frequency bandwidth ratio, which sets the frequency components according to the relationship between the pitches of notes, which is different from the traditional DFT method with equal frequency intervals, and the comparison between CQT and DFT is shown in Table 1.CQT provides a more reasonable spectral representation of musical signals in the frequency domain, especially in the problems related to the pitch frequency of the music, so as to better describe the harmonic characteristics of the musical sound without conflicting with the music theory. representation to better characterize the harmonic properties of musical sounds without conflicting with the expression.

Table 1. A comparison between CQT and DFT

Parameter variable	CQT	DFT
Frequency component	Exponential form distribution	Linear distribution
Resolution	δ_f Variable	Sampling rate window length ratio, constant
Window length	N_k Variable	Fixed
Window length	Q, constant	Variable

The chromaticity of each frame is defined as shown in Equation (10), where $b \in [1, \beta]$ denotes the number of chromaticity points and M denotes the number of octaves within the Q spectrum. Namely:

$$Chroma(b) = \sum_{m=0}^{M} \left| X^{cq}(b + m\beta) \right| \tag{10}$$

The resulting data is processed by tuning operation, correction of deviation data, low-pass filter smoothing, and then combined with endpoint detection to find out the average value of the chromaticity profile within each note fragment, and finally summed up within each semitone to convert the 36-dimensional data into 12 dimensions, so as to get the chromaticity eigenvectors of each single note or chord.

Calculate the Euclidean distance between the sequence of performance audio P eigenvalues $chro_P(i), i = 1 \cdots N$ and the sequence of score audio S eigenvalues $chro_S(j), j = 1 \cdots M$ to get the similarity matrix $SM(i, j)$, and then compute the cumulative matrix D by adding the weighting factors as shown in Equation (11) shown:

$$D(i, j) = SM(i, j) + \min \begin{Bmatrix} \omega_a * D(i, j-1) \\ \omega_a * D(i-1, j) \\ \omega_b * D(i-1, j-1) \end{Bmatrix} \tag{11}$$

where $D(1,1) = SM(1,1)$, the best experimental results are obtained by this system by taking the weights $\omega_a = 1$ and $\omega_b = 1.2$, and the path itself is obtained by $D(i, j)$.

The comparative local mapping $\mu : [1:N] \to [1:M]$ of S and P is monotonically increasing, ensuring the irreversibility of the alignment time. Ultimately, the comparison between the performance audio and the score audio is obtained by determining the DTW path, which determines the distance paths $W = W_1, \cdots, W_m$, with each W_k corresponding to an ordered pair (i_k, j_k) implying that S_i and P_j are aligned.

Extracting feature values based on endpoint detection is crucial to the error detection function, and in this paper, we analyze the correspondence between each note point of the audio of the performance (which includes chords and single notes) and the music score to determine whether the performance is correct or not.

A similarity matrix threshold σ is set, and the ordered pairs corresponding to the path W_n are (i_n, j_n) , and the j th note of the score in the performance is considered to be wrongly played when $SM(i, j) < \sigma$.

The path direction shows whether there are overplayed or missed notes. When $N = M$, and S corresponds to P , there is no overplay or omission (this does not exclude the possibility of a mistake). When $M > N$, the path W_{n+1} corresponds to the ordered pair $(i+1_{n+1}, j_{n+1})$, then there is no score corresponding to the $i+1$ th note segment in the performance audio, and the $i+1$ th note segment is an overplayed single note or chord. Similarly, when $N > M$, the path W_{n+1} corresponds to the ordered pair $(i_{n+1}, j+1_{n+1})$, indicating that the $j+1$ th note segment in the score is missed.

2.2. Psychological development in voice education

Although the specific voice training in vocal education given above can effectively improve the students' vocal theory and skills, it is difficult to realize the psychological quality training in vocal education without a detailed plan for psychological quality training. In this regard, on the basis of the previous article, to carry out voice education in the psychological quality training work, psychological quality is the individual students in the behavioral performance of the psychological state behind, different individual students, because of the innate character conditions and the growth of the environment conditions are different, the psychological quality of a great difference. Therefore, if teachers want to reasonably cultivate students to form good psychological quality, they must teach students according to their learning needs. According to the performance of the students during the sound training, choose the appropriate way to carry out psychological education and guidance, to ensure that the students build up self-confidence in playing, to overcome the psychological pressure of playing and bad emotions.

2.2.1. Definition of symbols

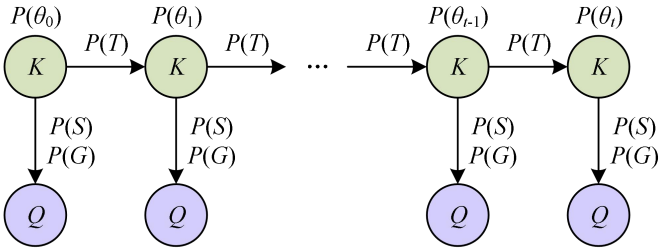
The symbols in the recommendation model and the corresponding descriptions are shown in Table 2, and the following design is made in combination with the actual situation of vocal music education. The recommended method for mental quality training in vocal education uses a single-valued neutrosophic set to portray the students' state of knowledge on knowledge points in vocal training in three aspects: mastery level, degree of misunderstanding, and uncertainty. A triad $A_{pv} =< t_{pv}, f_{pv}, i_{pv} >$ is used to characterize the student's p knowledge state on knowledge point v , and the three parameters characterize the student's p mastery level, misinterpretation level, and uncertainty on knowledge point v , respectively. The set of student's tuples on all k knowledge points is denoted as $A_p = \{A_{p1}, A_{p2}, \cdots, A_{pk}\}$.

Table 2. Symbol description table

Symbol	Description
A_{pv}	The knowledge status of student p at vocal training point v
P	Students gather
V	A collection of theoretical knowledge points in vocal music training
t_{pv}	Student p's mastery of knowledge point v
f_{pv}	The extent to which student p misunderstands knowledge point v
i_{pv}	Student p's uncertainty e regarding knowledge point v
$R'_{pv} = \{r'_{pv1}, r'_{pv2}, \cdots, r'_{pvn}\}$	The set of practice scores of student p on knowledge point v
$r'_{pve}, e = 1, 2, \cdots, n$	Student p scored e in the e-th training on knowledge point v

2.2.2. Knowledge tracking model

Referring to the current mainstream methods SB-CF (student-based collaborative filtering resource recommendation), EB-CF (content-based collaborative filtering resource recommendation), and KCP-ER (personalized exercise recommendation method based on the prediction of knowledge concepts), we finally chose the Bayesian Knowledge Tracking-based Recommendation Model (BKT-ER).BKT is a representative model of the Bayesian Knowledge Tracking model, which models the process of students' learning state transfer by means of a hidden Markov model. Markov model to model the process of transferring students' learning state. BKT is a model for modeling knowledge points, that is, one knowledge point corresponds to one model. It divides the variables into two sets: the hidden state set $S = \{0,1\}$ and the observable variable set $V = \{0,1\}$. The hidden state set represents the cognitive state of the student, while the observable variable set represents the student's dynamic practice performance. Both sets are described using 0-1 variables, where 0 indicates the "unmastered" state in the hidden state set and the "error" state in the observable variable set, and 1 indicates the "mastered" state in the hidden state set and the "correct" state in the observable variable set. The change of students' cognitive state of the knowledge points during the practice is the state transfer of HMM, and the schematic of BKT model is shown in Fig. 6.



$P(\theta_0)$: Probability of initially mastering a concept

$P(T)$: Transition probability from not mastering to mastering a concept

$P(G)$: Probability of answering a question correctly when the concept is not mastered

$P(S)$: Probability of answering a question incorrectly when the concept is mastered

Q: Question observation results

K: Question concept

Figure 6. BKT model diagram

The BKT assumes that once a student's mastery status changes from “not mastered” to “mastered”, it will not be reversed, and the core idea is to estimate the student's mastery of the knowledge point by taking into account the error rate and guessing rate. The model estimates four parameters: $P(S)$ is the probability of answering a question wrongly when the student has mastered the knowledge point, i.e., the error rate. $P(G)$ is the probability of answering the question correctly when the student is in the state of not mastering the knowledge point, i.e., the guessing rate. $P(T)$ is the probability that a

student will change from “not mastered” to “mastered” after practicing, i.e., the transfer probability. $P(\theta_0)$ is the initial probability that the student has mastered the current knowledge point before starting the exercise.BKT updates the parameters by the following formula:

$$P(\theta_t \mid Q=1)=\frac{P(\theta_{t-1})(1-P(S))}{P(\theta_{t-1})(1-P(S))+(1-P(\theta_{t-1}))P(G)} \tag{12}$$

$$P(\theta_t \mid Q=0)=\frac{P(\theta_{t-1})P(S)}{P(\theta_{t-1})P(S)+(1-P(\theta_{t-1}))(1-P(G))} \tag{13}$$

$$P(\theta_t)=P(\theta_{t-1})+(1-P(\theta_{t-1}))P(T) \tag{14}$$

$$P(Q_{t+1}=1)=P(\theta_t)(1-P(S))+(1-P(\theta_t))P(G) \tag{15}$$

$P(\theta_t)$ is the probability of students' mastery of the knowledge point at time step t, and Q is the observation sequence, i.e., the results of students' answers to the test questions, $Q\in\{0,1\}$. Equation (12) and Equation (13) estimate the probability of mastering the knowledge point at time step t based on the student's training results, when $Q=1$, i.e., when the answer is correct, it is calculated according to Equation (12), otherwise it is calculated by Equation (13). The conditional probability of knowledge point mastery at time step t is updated by Eq. (14). After obtaining the optimal solution of the parameters, the probability of the student's correct response at the next time step $t+1$ is calculated by Eq. (15). Although the BKT is practiced by students dynamically to update the parameters, the model assumes that the four parameters are the same for all students, ignoring the differences in characteristics between each student.

2.2.3. Modeling the state of vocal training

The model takes as input a collection of vocal training data for n students $X=\{X_1,X_2,\cdots,X_m\}$, where $X_p=\{x_1,x_2,\cdots,x_t\}$ is the practice of student p. data sequence. For the student's interaction records, this paper uses feature intersection to characterize them. The answer result s_i of each interaction record is represented by a uniquely hot encoding of length 2. The encoding length of vocal training e_i is related to the number of knowledge points k. After feature crossover, each interaction record is characterized by a vector of dimension $2k$, with the first k bits denoting knowledge points and the second k bits denoting knowledge point errors, when the n th knowledge point is correct the n th position is 1 and the rest of the bits are 0. When the n th knowledge point is answered incorrectly the $k+n$ th position is 1 and the rest of the bits are 0. Thus the inputs to the model are finally represented as One-Hot vector $x_i=(e_i,s_i)$ with $x_i\in R^{2k}$ containing the index of knowledge points and the answer results. Finally Att goes through a fully connected layer and sigmoid activation function to obtain the final output. The dimension of the output vector is the same as the number of knowledge points, and each bit represents the probability that the student will answer the knowledge point correctly at the next moment. I.e:

$$y_t=sigmoid\left(W_{out}\cdot Att\right) \tag{16}$$

The variables in the model, including the embedding matrices A, B, and the mapping matrices of the multi-head self-attention mechanism are obtained through training, and our training goal is to obtain the parameters of the model by minimizing the cross-entropy between the model's predicted value y_t and the true answer situation s_t:

$$L=-\sum_t\left(s_t\log y_t+(1-s_t)\log(1-y_t)\right) \tag{17}$$

The “level of misunderstanding” of a knowledge point refers to the degree of error or bias in a student's understanding and application of a knowledge point. In order to measure the degree of misunderstanding of a student p on a knowledge point v, the methodology introduces the definition of score loss rate, which takes into account the scores of individual knowledge points as well as the overall scores of the students, where the lower the score of a student on a certain knowledge point, the greater the misunderstanding of the student's understanding of that knowledge point may be. The calculations are shown in equations (18) to (20). To wit:

$$\beta_{pv} = \sum_{e=1}^{|R'_{pv}|} 1 - r'_{pve} \tag{18}$$

$$\gamma_v = \max(\beta_{pv}), p \in P \tag{19}$$

$$f_{pv} = \frac{\beta_{pv}}{\gamma_v} \tag{20}$$

where r'_{pve} is the score of student p on knowledge point v on the e th exercise, β_{pv} is the total score loss of student p on knowledge point v, and γ_v is the maximum value of score loss on knowledge point v in the set of students P.

In the measure of knowledge point mastery, entropy reflects the degree of discretization and ambiguity in the student's ability to understand and apply the knowledge point. The expectation, entropy and hyperentropy of student P on knowledge point v are calculated in the following manner:

$$\hat{E}_x^{pv} = \frac{\sum_{e=1}^{|R'_{pv}|} r'_{pve}}{\left|R'_{pv}\right|} \tag{21}$$

$$\hat{E}_n^{pv} = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{\left|R'_{pv}\right|} \sum_{e=1}^{|R'_{pv}|} \left| r'_{pve} - \hat{E}_x^{pv} \right| \tag{22}$$

$$\hat{H}_e^{pv} = \sqrt{\left| \frac{1}{\left|R'_{pv}\right| - 1} \sum_{e=1}^{|R'_{pv}|} (r'_{pve} - \hat{E}_x^{pv})^2 - (\hat{E}_n^{pv})^2 \right|} \tag{23}$$

where $\hat{E}_x^{pv}$, $\hat{E}_n^{pv}$, and $\hat{H}_e^{pv}$ are the expected value, entropy, and superentropy, respectively, of the student's p scores on knowledge point v. The uncertainty of the knowledge point i_{pv} is given by:

$$i_{pv} = \hat{E}_n^{pv} + \lambda \hat{H}_e^{pv} \tag{24}$$

2.2.4. Specific processes

After calculating the students' knowledge point mastery, misinterpretation level and uncertainty, we further adjusted the students' overall level of psychological quality. Firstly, the overall mastery of knowledge points t_p, the overall misunderstanding level f_p and the overall uncertainty i_p of students p during all the sound training were calculated by using equations (25) to (27). To wit:

$$t_p = \sqrt[|V|]{\prod_{v=1}^{|V|} t_{pv}} \tag{25}$$

$$f_p = \sqrt[|V|]{\prod_{v=1}^{|V|} f_{pv}} \tag{26}$$

$$i_p = \sqrt[|V|]{\prod_{v=1}^{|V|} i_{pv}} \tag{27}$$

where V is the set of knowledge points, t_{pv} is the degree of mastery of student p on knowledge point v, f_{pv} is the degree of misinterpretation of student p on knowledge point v, and i_{pv} is the uncertainty of student p on knowledge point v.

The larger values of the degree of misunderstanding and uncertainty indicate that the degree of students' precise mastery of the knowledge points is smaller. The comprehensive score of students' vocal training is evaluated by the calculation of formula (28), and finally the overall level of students' psychological quality is adjusted in accordance with (29), so as to facilitate the next step of recommending the matching of students with the appropriate music resources. The value of the parameter ω is discussed in the experimental section. To wit:

$$lev_p = \omega \times t_p + (1-\omega) \times (1 - \frac{f_p + i_p}{2}), \quad \omega \in [0,1] \tag{28}$$

$$Level_p = \begin{cases} Level_p + 1, & lev_p \geq 0.8 \\ Level_p, & 0.6 \leq lev_p < 0.8 \\ Level_p - 1, & lev_p < 0.6 \end{cases} \tag{29}$$

where lev_p is the student's p single diagnostic assessment score and $Level_p$ is the adjusted student level. The specific strategy flow of the recommendation model is as follows:

(1) Call the knowledge tracking model, the input parameters are the student's historical training record AnsList and the training difficulty difficultList, sound training duration List, to get the student's knowledge point state matrix knowledge State.

(2) Calculate the degree of misinterpretation and uncertainty of the knowledge point with the student score matrix score, and get the set A_p.

(3) Evaluate the overall composite score lev_p of students' mental quality according to the set A_p and adjust the mental quality level $Level_p$.

(4) Select music resources e in the test bank whose sound training difficulty level matches the student's level $Level_p$ and whose exposure meets the constraints, and recommend them to the students.

2.3. Integration of both applications

Vocal education is a kind of fusion of music and literature, a kind of communication between voice and emotion. Vocal education stems from the inner emotional experience of music, the sublimation of the connotation of the lyrics, for the scientific use of vocal skills, voice training and the cultivation of mental quality ability are the key teaching links and important means to implement effective vocal teaching.

Through in-depth mining and processing of relevant data, the improved DTW algorithm can accurately identify the problems of students in the singing process and generate detailed assessment reports, which not only contain objective scores on students' singing skills, but also provide targeted improvement strategies. For example, for pitch problems, the AI can specifically point out the notes or intervals that the student needs to strengthen in particular, significantly enhancing his or her vocal training.

The BKT-ER model can also intelligently recommend appropriate practice repertoire and training methods based on the student's performance during the voice training period, as well as selecting appropriate ways to conduct psychoeducation and guidance, with a view to enhancing the cultivation of the students' psychological quality in the vocal education of colleges and universities. The application of artificial intelligence in the professional education of vocal music can effectively promote the voice training and psychological quality cultivation, and help teachers better understand the learning situation of students, so as to timely adjust the teaching strategy and improve the teaching method.

3. Example Exploratory Analysis

3.1. Analysis of voice training in vocal music education

3.1.1. Base note extraction analysis

Basic frequency extraction experiments in a pure music environment and in a noisy language environment were set up. The experimental data in the pure music environment were obtained from the MIR-1K dataset provided by Hsu. The MIR-1K dataset contains 1000 Mandarin song fragments containing both male and female audio, and these 1000 16KHz sampled song fragments are in stereo, which can be read directly from their right channel in pure singing . Noise signals are added to the original data in MIR-1K at a signal-to-noise ratio of 0 dB to obtain the test audio for fundamental frequency extraction experiments in noisy environments, and the noise signals are taken from the Noise-92 dataset. Two main performance indicators are taken to judge the accuracy of fundamental frequency extraction in the experiment, which are:

FPE: denotes the percentage of erroneous audio frames with fundamental frequency extraction error less than 20% to the total number of frames.

FFE: indicates the percentage of fundamental frequency extraction error, i.e., the percentage of the number of erroneous frames to the total number of frames, and FFE can also be used as the only evaluation criterion of fundamental frequency extraction performance.

Table 3 shows the experimental results of fundamental frequency extraction in a pure music environment, where PPROC, STACF, ZCR, and SHS are used as the four compared fundamental frequency extraction algorithms. Table 4 shows the fundamental frequency extraction experiments under noise permutation. From Table 3 and Table 4, it can be seen that the AMDF algorithm proposed in this paper has the highest base frequency extraction accuracy compared to several comparison algorithms, and the base frequency extraction error is minimized compared to FFE, but the FFE reduction is limited. In the noise environment, compared to several comparison algorithms, AMDF algorithm base frequency extraction error is minimized than FFE, especially for female voice, base frequency extraction error rate is reduced by 23.65% at least. The experimental results can show that the AMDF algorithm improves the fundamental frequency extraction accuracy and enhances the noise robustness of the algorithm.

Table 3. Fundamental frequency extraction experiment in a noise-free environment

Project	Female		Male	
	FPE	**FFE**	**FPE**	**FFE**
STACF	2.925	8.033	2.811	9.814
ZCR	2.735	9.134	3.036	10.051
PPROC	1.934	7.815	2.805	11.224
SHS	1.925	7.805	2.815	9.871
AMDF	1.833	7.482	2.716	9.144

Table 4. Fundamental frequency extraction experiment in a noisy environment

Project	Female		Male	
	FPE	**FFE**	**FPE**	**FFE**
STACF	3.022	21.051	2.915	25.117
ZCR	2.724	26.032	3.407	29.722
PPROC	2.516	29.153	3.137	31.493
SHS	2.733	27.081	3.304	25.088
AMDF	2.616	16.072	3.432	22.073

3.1.2. Base tone smoothing analysis

On the basis of the above, the fundamental tone smoothing analysis was carried out, and other indexes were kept unchanged. The results of fundamental smoothing under pure music environment are shown in Table 5, and the results of fundamental smoothing under noise permutation are shown in Table 6. Based on the size of the data in the table, it can be seen that compared with using the median smoothing algorithm and linear smoothing algorithm alone, most of the interfering information is eliminated under the mixed effect of the two, which ensures the feasibility of voice training in vocal music education.

Table 5. Pitch smoothing results in a pure speech environment

Project	Female		Male	
	FPE	**FFE**	**FPE**	**FFE**
Linear smoothing algorithm	2.876	9.805	3.499	9.219
Median smoothing algorithm	2.667	9.291	3.243	8.778
Mix the two	2.613	6.442	2.241	6.474

Table 6. The result of pitch smoothing with noise substitution

Project	Female		Male	
	FPE	**FFE**	**FPE**	**FFE**
Linear smoothing algorithm	4.862	28.875	5	25.913
Median smoothing algorithm	4.567	27.017	3.355	25.446
Mix the two	4.481	21.175	3.072	24.685

3.1.3. Feature Matching Analysis

First, the sequences are set to be X and Y. The subsequence matching experiment is performed in the ideal state. Where the feature sequence X is set as a subsequence of feature Y, X=Y(150:500). Comparison of DTW algorithm and improved algorithm sequence matching results. Figure 7 shows the comparison results of subsequence matching. The dashed line in the figure indicates the DTW matching results and the solid line indicates the improved algorithm matching results. The traditional DTW algorithm enforces the matching start point and matching end point, and introduces the offset cost at the front and back ends of the matching path, and the final matching result includes the offset cost at the front and back ends in addition to the actual matching result (the distance value is 0), and the distance result of the DTW algorithm is not 0. The improved algorithm introduces the end-point detection mechanism, and there is no enforced restriction on the matching end point (the matching end point is the first point of sequence Y 492th point of Y sequence and 357th point of X sequence), the improved algorithm realizes subsequence matching, and the distance result is 0. The improved algorithm, compared with the original DTW algorithm, does relax the restriction of the matching start point and end point, and objectively realizes subsequence similarity comparison.

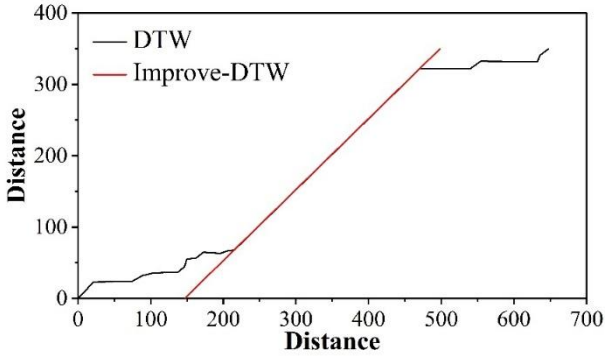


Figure 7. The comparison result of subsequence matching

3.1.4. Example analysis of vocal education

In the vocal training, 10 singers with different singing standards were selected to sing the same song, labeled as Singers A, B, C, D, E, F, G, H, J, and K, according to their singing level from highest to lowest. At the same time, an eight-member scoring panel was formed to score the songs sung by each singer, and the scoring was done entirely according to their subjective preferences, giving an assessment score of the degree of melodic mastery of the entire song sung by the singers, and the score was based on a percentage system. The scores were labeled as one, two, three, four, five, six, seven, and eight. The experiment will compare the actual scoring results with the scoring results based on the improved DTW feature matching, and Table 7 demonstrates the experimental results. The data performance in the table shows that the average difference between the scoring results of the improved DTW feature matching and the actual scoring is 0.05~0.63, and its accuracy rate is kept above 90%, which indicates that the improved DTW algorithm for vocal feature matching can help the students to deeply recognize their self-errors in the training process, which in turn helps the students to grasp the situation of their voice training.

Table 7. Experimental results

Project		A	B	C	D	E	F	G	H	I	J
Actual scoring result	1	88	81	63	71	83	73	53	68	48	86
	2	67	45	75	52	68	72	71	48	55	59
	3	79	71	82	49	56	80	64	89	67	77
	4	43	51	60	81	54	63	63	68	45	76
	5	51	53	63	69	63	44	61	61	64	82
	6	58	63	66	90	57	59	44	50	46	48
	7	46	82	73	60	50	77	64	52	53	80
	8	56	81	68	79	61	86	80	44	65	61
	Mean	61.00	65.88	68.75	68.88	61.50	69.25	62.50	60.00	55.38	71.13
DTW		61.27	65.93	68.87	68.94	62.07	69.88	62.67	60.29	55.46	71.18

3.2. Analysis of psychological quality cultivation in vocal music education

The music recommendation algorithm based on knowledge tracking aims to deeply understand

learners' real needs and learning goals, and then provide appropriate music resources according to their knowledge level to ensure that students build up self-confidence in playing, overcome the psychological pressure and bad emotions in playing, and then promote the development of psychological quality cultivation work in vocal music education in colleges and universities. In this subsection, we will evaluate the effectiveness of the above proposed method through experiments.

3.2.1. Assessment of indicators

The work of psychological quality cultivation in vocal education is essentially to recommend appropriate teaching resources based on the students' performance during voice training, so that they can establish a good level of psychological quality. Therefore, this paper draws on the commonly used evaluation indexes of accuracy for evaluation, in addition to the introduction of novelty and diversity, and finally establishes the overall psychological quality cultivation effect indexes to ensure that the results of the study are more convincing.

(1) Accuracy is a measure of whether or not the recommended music resources are given to the students that are suitable for their level of psychological quality, as shown in formula (30), the greater the accuracy, the more the recommended music resources are in line with the learners' level of psychological quality. Namely:

$$Accuracy\left(L^*\right)=\frac{\sum_{i=1}^{M}\left(1-\left|\delta-D_{q_i(k)}\right|\right)}{\left|M\right|} \tag{30}$$

(2) Novelty

Novelty (Novelty) is an important indicator of how good a recommended music resource is, and when the recommended music resource contains more knowledge points that students have not been exposed to or have not mastered before, the higher its novelty is. Namely:

$$Novelty\left(L^*\right)=\frac{1}{\left|M\right|}\sum_{i=1}^{M}\left(1-\cos\left(H\left(q_i\left(k\right)\right),H\left(q_i^{v}\left(k\right)\right)\right)\right) \tag{31}$$

(3) Diversity

Is an important indicator for evaluating the differences between recommended music resources. The diversity value can help whether the recommendation model can cover several different knowledge areas when providing music resources, and high diversity means that the recommended music resources cover a variety of different knowledge points, which can stimulate students' learning interest and motivate them to study in depth in multiple areas. The calculation formula is shown in Equation (32):

$$Diversity\left(L^*\right)=\frac{\sum_{q_i(k)\in L^*}\sum_{q_j(k)\in L^*}\left(1-\mathrm{si}\,m\left(q_i\left(k\right),q_j\left(k\right)\right)\right)}{\left|M\right|\left|M-1\right|} \tag{32}$$

(4) Overall psychological quality cultivation effect

On the basis of accuracy, novelty, and diversity, the fixed adjustment factors of 0.4, 0.3, and 0.3 were respectively assigned to finally obtain the overall psychological quality cultivation effect indicators. The specifics are as follows:

$$Total=0.4*Accuracy+0.3*Novelty+0.3*Diversity \tag{33}$$

3.2.2. Results and analysis

In order to further validate the effectiveness of the BKT-ER proposed in this paper, this paper compares it with some classical methods and the latest existing methods.Fig. 8 shows the results of the analysis of the resource recommendation in psychological quality training, where (a) ~ (d) are accuracy, novelty, diversity, and psychological quality training effect, respectively.The closer the Mean value is to 1 in the figure, the better the performance of the model is indicated. In addition, Std.D indicates the standard deviation of each index, which reflects the stability of the recommendation effect. The benchmark algorithms are as follows:

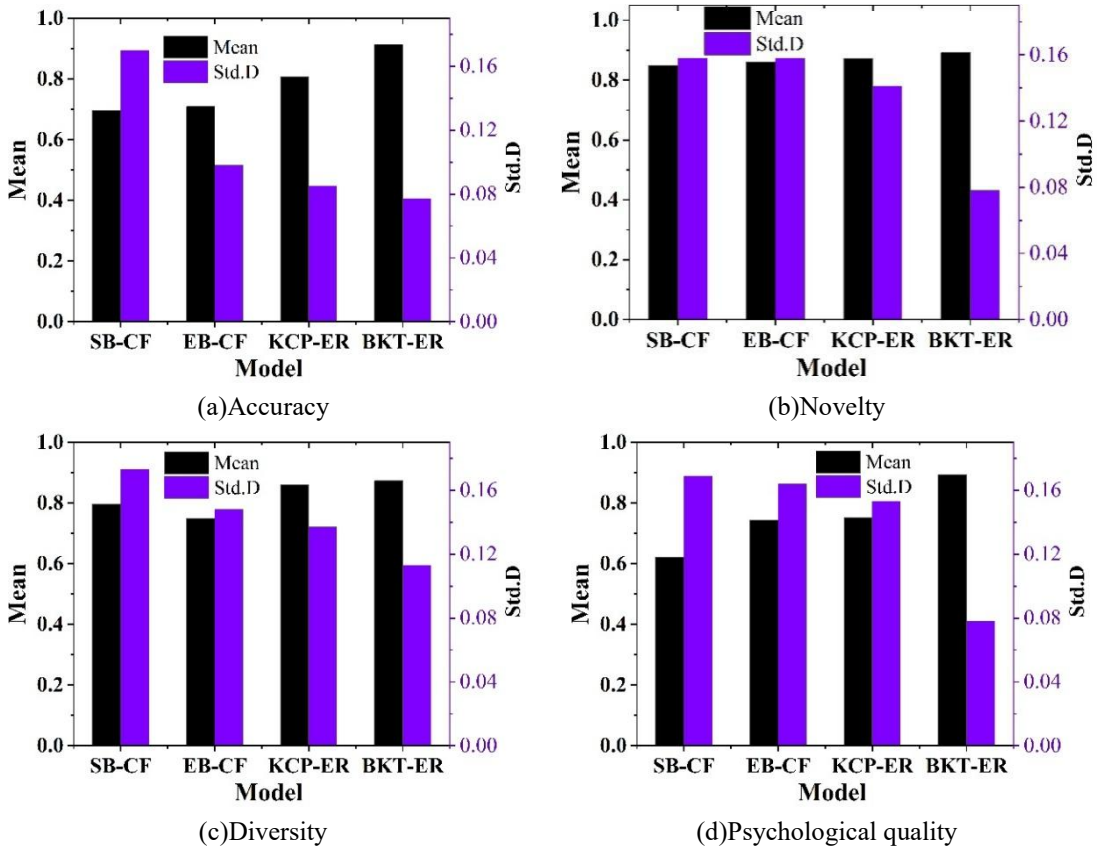
(1) SB-CF: It is student-based collaborative filtering resource recommendation. The model is using the classical student-based collaborative filtering algorithm to recommend exercises for students.

(2) EB-CF: is content-based collaborative filtering resource recommendation, the model adopts the

idea of item-based collaborative filtering in recommender systems to recommend exercises for students.

(3) KCP-ER: a personalized exercise recommendation method based on knowledge concept prediction is proposed.

Based on the data performance in the figure, it can be seen that the BKT-ER model shows significant effectiveness in terms of accuracy, novelty, diversity, and mental quality development effect conducted on the dataset. This means that the BKT-ER model is able to recommend appropriate teaching resources based on students' performance during voice training, thus better meeting the requirements of psychological quality training work in college vocal education.



(a)Accuracy (b)Novelty

(c)Diversity (d)Psychological quality

Figure 8. Analysis of Resource Recommendation in Psychological Quality Cultivation

3.3. Analysis of the effect of the integration of the two applications

Randomly selected 50 students as the research object, according to the above application program into the college vocal education, using vocal education test scale for students to test, in order to obtain the two integration of the application of the effect of analysis of the research data, the scale is divided into the theoretical knowledge, basic skills, emotional regulation, psychological cognitive ability, and the scale has excellent reliability performance, to ensure that the results of the study is more in line with the actual situation of vocal music education. The actual situation, the results of the fusion of the two application effects are analyzed as shown in Table 8. Based on the size of the data in the table, it can be seen that the mean values of basic skills, emotional regulation ability and psychological cognitive ability are 3.606, 3.846, 3.796, 3.894, respectively, and the corresponding standard deviations are 0.881, 0.635, 0.713, 0.725, in short, the mean values of all dimensions are kept above 3.5, which indicates that under the dual role of voice training and psychological quality cultivation, students' basic skills in vocal education can be improved. Under the dual role of voice training and psychological quality cultivation, students' basic skills, emotional regulation ability, and psychological cognitive ability in vocal education are further improved, fully verifying the usability of the research program in this paper.

Table 8. Analysis results of the integrated application effect of the two

Project	N	Min	Max	Mean	SD

Theoretical knowledge	50	1	5	3.606	0.881
Basic skills	50	1	5	3.846	0.635
Emotional regulation ability	50	1	5	3.796	0.713
Psychological cognitive ability	50	1	5	3.894	0.725

4. Conclusion

Combined with the actual situation of vocal education in colleges and universities at present, this paper establishes a vocal training assessment program based on improved DTW, a psychological quality development program based on BKT-ER, and brakes the integration program of the two, respectively, with a view to improving the level of development of vocal education in colleges and universities.

(1) The difference between the scoring results of improved DTW feature matching and the actual scoring is less than 0.63, and the corresponding accuracy rate is greater than 90%, i.e., it shows that using the improved DTW algorithm for vocal feature matching can help students deeply recognize their own mistakes during the training process, thus improving the quality of students' voice training in vocal music education.

(2) BKT-ER shows excellent application value in terms of accuracy, novelty, diversity, and psychological quality cultivation effect, and the Mean value of each index is greater than 0.8, which indicates that the BKT-ER model is able to provide appropriate teaching resources based on the students' performance during the vocal training period, which in turn meets the goal of psychological quality cultivation in vocal education in colleges and universities.

(3) Under the effect of the fusion of the two programs, the mean value of students' basic skills, emotional regulation ability, and psychological cognitive ability in vocal education is greater than 3.5, i.e., it proves that the fusion of the two programs is effective in the practical application of vocal education.

About the Author

Shifang Yang, female, born in November 1984, postgraduate, lecturer. She graduated from the School of Music, Southwest University, majoring in Musicology. She is currently working at Chongqing Youth Vocational & Technical College, with her main research interests in vocal music performance and teaching research. She mainly teaches courses including Basic Vocal Music, Music Theory, and Solfège & Ear Training.

References

1. Du, Q. (2024). A systematic approach to innovative strategies for vocal instruction in higher education: Enhancing student performance. Pacific International Journal, 7(5), 68-73.
2. Yin, W. (2024). Innovations and practical exploration of vocal music teaching models in vocational colleges. Journal of Modern Educational Theory and Practice, 1(2).
3. Jing, W. (2020). New Exploration of Vocal Music Education in Colleges. Frontiers in Educational Research, 3(1).
4. Paulmann, S., & Weinstein, N. (2025). Motivating tones to enhance education: The effects of vocal awareness on teachers' voices. British Journal of Educational Psychology, 95(2), 551-564.
5. Zhou, L. (2023). Cultivation of artistic expression in college music and vocal music teaching. Art and Performance Letters, 4(12), 43-49.
6. Pabon, P., Stallinga, R., Södersten, M., & Ternström, S. (2014). Effects on vocal range and voice quality of singing voice training: the classically trained female voice. Journal of Voice, 28(1), 36-51.
7. Thompson, D. L. (2025). Vocal Development Using Student-Centered. Student-Centered Voice Pedagogy: Working with Students Toward Developing Artistry, Authenticity, and Autonomy, 152.
8. Angadi, V., Croake, D., & Stemple, J. (2019). Effects of vocal function exercises: a systematic review. Journal of Voice, 33(1), 124-e13.
9. Rebro, I. V., Mustafina, D. A., Rakhmankulova, G. A., Mashikhina, T. P., & Aleksandrina, A. Y. (2024). The relationship of abilities and psychological attitudes in the process of carrying out activities. Review of pedagogical research, 6(2), 15-23.

10. Duan, G., Zeng, J., & Ji, H. (2021). The Importance of Psychological Analysis to Vocal Music Singing Teaching. Psychiatria Danubina, 33(suppl 7), 315-317.
11. Hendricks, K. S. (2016). The sources of self-efficacy: Educational research and implications for music. Update: Applications of Research in Music Education, 35(1), 32-38.
12. Zhou, N. (2022). A study on the teaching of vocal music in colleges and universities from the perspective of aesthetic education. Advances in Educational Technology and Psychology, 6(1), 25-28.
13. Rauduvaite, A., & Yao, Z. (2020, May). Prospective music teacher training: factors contributing to creation of positive state in the process of vocal education. In Rural Environment Education. Personality (REEP). In Proceedings of the 13th International Scientific Conference, Jelgava, Latvia (pp. 8-9).
14. Zhang, X. (2023). An Analysis of the Method of Combining Vocal Teaching and Music Psychology. The Educational Review, USA, 7(7).
15. Pan, Y. (2021). Research on the role of singing psychological quality in vocal music teaching and performance. Psychiatria Danubina, 33(suppl 8), 495-496.
16. Ye, Y. (2020). Application of positive psychology in classroom teaching of vocal music. Revista Argentina de Clínica Psicológica, 29(1), 353.
17. Yang, C. (2015, October). Research on Function of Psychological Factor in Vocal Music Teaching. In International Conference on Education, Management and Information Technology (pp. 163-167). Atlantis Press.
18. Wu, J. (2022). RESEARCH ON THE INFLUENCE OF VOCAL MUSIC TEACHING MODE BASED ON EDUCATIONAL PSYCHOLOGY ON STUDENTS'STAGE ANXIETY. Psychiatria Danubina, 34(suppl 4), 21-21.
19. Hongkun, Z. (2024). Impact of Music Style Vocal Training on Psychological Performance Among College Students at Shandong University of Arts, China. International Journal of Academic Research in Business and Social Sciences, 14(7), 554-564.
20. Wang, R. (2018, April). The Influence of Mental State on Vocal Music Learning. In Proceedings of the 2018 1st International Conference on Internet and e-Business (pp. 375-377).
21. Zixuan, L. (2021). Research on the current situation of Chinese national vocal music education from the perspective of educational psychology. Frontiers in Art Research, 3(6), 48-58.
22. Wang, X. (2022). Psychology education reform and quality cultivation of college music major from the perspective of entrepreneurship education. Frontiers in psychology, 13, 843692.
23. Dong, J. (2022). RESEARCH ON THE CULTIVATION OF STUDENTS'SINGING PSYCHOLOGICAL QUALITY IN VOCAL MUSIC EDUCATION. Psychiatria Danubina, 34(suppl 4), 506-506.
24. Wu, J. (2024). The Importance of Building Students' Good Singing Psychology. International Journal of New Developments in Education, 6(10).
25. Wang, R. (2023). The influence of music education combined with sports psychology on students' anxiety. Revista de Psicología del Deporte (Journal of Sport Psychology), 32(2), 305-312.
26. Lang, Q. (2016, July). Application of psychological regulation ability in vocal teaching. In 2016 5th International Conference on Social Science, Education and Humanities Research (SSEHR 2016) (pp. 103-107). Atlantis Press.