Hindawi
Mobile Information Systems
Volume 2022, Article ID 9747342, 11 pages
https://doi.org/10.1155/2022/9747342

Research Article

Intelligent Analysis of Music Education Singing Skills Based on Music Waveform Feature Extraction

Rong Li

Xinyang Vocational and Technical College, Xinyang 464000, China

Correspondence should be addressed to Rong Li; lirong2022@xyvtc.edu.cn

Received 7 April 2022; Accepted 4 May 2022; Published 31 May 2022

Academic Editor: Yajuan Tang

In order to improve the effect of intelligent analysis of singing skills in music education, this paper conducts an intelligent analysis of singing skills in music education with the support of music waveform feature extraction technology. Moreover, this paper uses the traditional IMM optimal waveform selection algorithm to solve the model and analyzes the tracking effect and the changes of transmission parameters. Compared with the fixed transmission parameters, the algorithm can effectively reduce the tracking error. From the experimental analysis, it can be seen that the intelligent analysis system for music education singing skills based on the extraction of music waveform features has good effects and can effectively promote the improvement of music singing skills and the improvement of music teaching effects.

1. Introduction

Music information visualization refers to the visual presentation of music information based on the principle of music acoustics to make it intuitive, so as to eliminate the disadvantages of the ambiguity of music sound in music education and music research and improve the accuracy and effectiveness of music information transmission.

According to basic music theory, the four elements of musical sound are divided into four points: pitch, sound intensity, timbre, and sound length. The combination of these four elements constitutes the colorful music we hear today. However, these four elements were not known to people at the beginning of the birth of music but were derived from the continuous practice and theoretical development of music, the development of basic sciences such as mathematics and physics, and the progress of science and technology. Moreover, the development of music practice and music theory is not only accompanied by changes in the "connotation" and "extension" of the concept of "music" in various eras, but also accompanied by changes in aesthetics and creative methods. These changes are not only influenced by social factors such as politics, religion, and culture, but also the technological and scientific progress behind them are also driving forces that cannot be ignored.

In the field of computer music, singing voice synthesis is a comprehensive application of multiple research works, which often involves many aspects of the field of speech synthesis, including endpoint detection and extraction of the underlying features of the human voice, especially the fundamental frequency and timbre features of the human voice extract [1]. Among them, endpoint detection, speech segmentation, fundamental frequency extraction, and other research work have been carried out earlier and have more research results, while the extraction of human voice color features is still an emerging research branch. Song synthesis technology is another new research hotspot in the field of music speech signal processing after speech recognition, music retrieval, and music recommendation. First, it can be applied to the singing synthesis of "virtual" singers; secondly, the application of singing synthesis to music singing education can reduce the recording of repetitive singing teaching materials, thereby reducing the waste of human and material resources [2].

There are four important terms that musicians use to describe sounds in music: length, intensity, pitch, and timbre. The listener can sort the sounds from high to low

according to the pitch or intensity of the sound, and the length of the sound depends on the duration of the sound of the object [3]. Timbre is used to describe the subjective auditory properties of a sound with a specific pitch and intensity, and there is no objective metric. Among them, pitch and timbre are important essential characteristics of sound, and many scholars have studied and explored the modification of pitch and timbre. This topic mainly discusses the pitch and timbre of speech, especially the pitch and timbre of the singing voice when singing songs [4]. Literature [5] has written a review to summarize the modification of the current speech pitch, in which the waveforms are similar and overlapped, and selecting the overlapping segment according to the waveform similarity to modify the pitch can keep other features of the music unaffected. Reference [6] proposes the superposition of pitch synchronization waveforms and finds superimposed waveforms according to the pitch period. Pitch modification within a small amplitude range can achieve good results. Reference [7] proposed to use the constant Q transform to perform time-frequency conversion and to modify the spectral pitch according to the characteristic curve in line with human hearing. Reference [8] proposed a pitch modification method based on the source-filter model and linear prediction coefficient, which separates the sound source information from the channel response and can perform pitch modification without affecting the channel response. There are many factors that affect timbre, including spectral envelope, formant [9], and spectral centroid [10], among which spectral envelope is the most direct feature that affects timbre. For example, the voice of the same person, even if the pitch is inconsistent, the timbre expressed will be different to a certain extent, but the spectral envelope will not change greatly. Reference [11] proposed a method that can modify the formants so that the timbre characteristics of a specific person can be changed by modifying the formants through technical processing. Most of the research on timbre analyzes the timbre characteristics of the audio signal as a whole. With the breakthrough of artificial intelligence and machine learning technology, data-driven audio feature analysis has become a research hotspot. Pitch conversion is an important technology in speech sound processing, and it is also the most basic and critical issue in singing harmony. Singing pitch conversion is a technique to adjust the raising or lowering of the singing pitch without changing the semantics and speech rate. Pitch conversion is widely used in song singing, music post-processing, KTV sound effects production, live video, and other scenarios. Vocal range refers to the range of pitches that a singer can emit according to his own physiological characteristics, and converting the pitch can expand the vocal range of the singer to a certain extent [12]. At present, the methods of pitch conversion are mainly divided into three categories: time-domain method, frequency-domain method, and parametric method. In the time-domain processing, resampling is the main method to achieve pitch conversion. In the frequency-domain processing, the frequency spectrum can be directly multiplied by the proportional coefficient to modify, while in the parametric method, the pitch is mainly based on the model characterization parameters of the sound, which modifies directly. However, the existing pitch conversion technology mainly focuses on the processing of ordinary speech signals, and when applied to singing voices, there are shortcomings such as unnatural conversion, drastic changes in timbre, and large distortion [13].

With the diversification and intelligent development of digital media technology, the combination of traditional media and emerging digital media in various fields has become more and more closely [14]. The transmission and exchange of information, on the one hand, is expanding from material media to digital media, and on the other hand, it is imperative to rapidly change from single media to multimedia, mixed media, and even integrated media [15].

With the transformation of media forms, as the most representative auditory art in digital media art, electronic music not only has a greater space for development in terms of creation and expression but also has a great impact on its artistic evaluation criteria. Some electronic music has been separated from the pure note system of traditional music and transformed into an independent music system with sound as the basic structure [16]. In an all-media environment, an art form that combines electronic music, visual art, and even digital intelligence technology will break the original creation and aesthetic rules, thus forming a new art form. In the field of research, the related research on the digital art creation mode and omnimedia expression of computer-led human intervention and collaborative creation is still in the exploratory stage. Under the background of increasingly intelligent, humanized, convenient, and interactive digital media, from the perspective of media, combined with the style characteristics, creation, performance, appreciation methods, and aesthetic system of electronic music, it is of great practical significance to conduct in-depth excavation and research on the "Music Creation and Performance System," establish a relatively complete system and the digital logic model, and provide theoretical and technical support for the Omnimedia development of intelligent electronic music creation and artistic expression in my country [17]. At the same time, the creative concepts and modes of intelligent electronic music are not limited to the creation of academic and experimental electronic music. In the future, it can also be applied to the creation of electronic dance music, ambient music, music installations, and other artistic works [18].

The volume is determined by the amplitude of the sound wave. The larger the amplitude, the higher the volume, and the smaller the amplitude, the lower the volume. If it is the same instrument, the greater the strength, the louder the volume, and vice versa. The volume change in music is the main root cause of musical expression and has a direct impact on musical expression. The difference in the overall volume will also make the same piece of music have different effects and give people different feelings. Each object has its own volume limit, which is usually determined by the object's size, material, and structural shape. The volume change also affects the timbre, which affects different objects to different degrees. For example, when playing a timpani with light intensity and heavy intensity, there is a big difference in

the timbre. When playing lightly, the pitch is more obvious, and the sound is round and full. It is similar to the bass of pizzicato, and the pitch is relatively less obvious when playing with heavy force, adding a lot of noise, which is closer to the sound of the bass drum, and the sound difference between heavy and light piano playing is not so obvious, but the sound of heavy playing is a little more noisy. Electronic audio devices rely on speaker vibrations to produce sound, and their overall volume depends on device performance and the voltage input to the speakers.

In this paper, with the support of music waveform feature extraction technology, the intelligent analysis of singing skills in music education is carried out, the recognition of singing skills in music education is improved, and the intelligent effect of modern music teaching is promoted.

2. Influence of Waveform on Tracking Effect of Music Recognition System

2.1. Influence of Waveform Parameters on Signal-to-Noise Ratio Received by Music Recognition System. According to the principle of music recognition system, when the tracking target is a point target and the environment is in the environment of Gaussian white noise with fixed noise power, the output SNR is related to the transmission energy, but not to the waveform parameters. In such a case, adjusting the parameters of the transmit waveform does not improve the SNR of the echo of the music recognition system.

The target's current environment contains noise interference and clutter interference is set. When the target to be tracked is an extended target, the output obtained by the music recognition system receiver through filtering is

$$y(t) = r(t)^* \left(s(t)^* h(t) + s(t)^* c(t) + J(t)\right). \tag{1}$$

In (1), $r(t)$ is the impulse response of the receiver, $h(t)$ is the impulse response of the target, $c(t)$ is the impulse response of the clutter, and $J(t)$ are other disturbances that the music recognition system experiences in its work. The result obtained by the filter output of the music recognition system receiver in (1) is decomposed, and the received signal components and noise components can be obtained as

$$\begin{aligned} y_s(t) &= r(t)^* s(t)^* h(t), \\ y_n(t) &= r(t)^* \left(s(t)^* c(t) + J(t)\right). \end{aligned} \tag{2}$$

In (2), $y_s(t) y_n(t)$ is the signal component and the noise component in the received echo, respectively. At time t_0, the output signal-to-noise ratio is

$$\begin{aligned} \mathrm{SNR}_{t_0} &= \frac{\left|y_s^2(t_0)\right|}{E\left[\left|y_s^2(t_0)\right|\right]}, \\ &= \frac{\left|\int_{-\infty}^{\infty} R(f) S(f) H(f) e^{j2\pi f_0}\, \mathrm{d}f\right|}{\int_{-\infty}^{\infty} |R(f)|^2 L(f) \mathrm{d}f}, \end{aligned} \tag{3}$$

where $|L(f)|^2 = P_c(f)|S(f)|^2 + P_J(f)$. Combined with Schwartz's inequality, the SNR expression can be obtained as follows:

$$\begin{aligned} \mathrm{SNR}_{t_0} &= \frac{\left|\int_{-\infty}^{\infty} R(f)\sqrt{L(f)}\,\left(S(f)H(f)/\sqrt{L(f)}\right) e^{j2\pi f_0} \mathrm{d}f\right|^2}{\int_{-\infty}^{\infty} L(f)|R(f)|^2 \mathrm{d}f} \\ &\le \int_{-\infty}^{\infty} \frac{|S(f)H(f)|^2}{L(f)} \mathrm{d}f. \end{aligned} \tag{4}$$

When the maximum value of SNR obtained by (4) is obtained, the form of the filter is

$$R(f) = \frac{kH(f)S(f)e^{j2\pi f_0}}{L(f)}. \tag{5}$$

To sum up, in the noise and clutter environment, when the target is an extended target, the parameters of the transmitted waveform of the music recognition system will have an impact on the signal-to-noise ratio of the received echoes.

2.2. Influence of Waveform Parameters on Measurement. When the music recognition system has detected the target, measure the influence of the waveform parameters on the measurement. During the calculation, in order to obtain the target delay-Doppler measurement error, the calculated target delay is used in the processing. The Cramer–Rao Lower Bound (CRLB) approximation of Doppler's maximum likelihood estimate serves as the target delay and measurement error for Doppler frequency. Using the parameter estimation theory, the Fisher matrix of the target delay-Doppler frequency can be obtained as follows:

$$J = \eta \begin{bmatrix} \overline{\omega^2} - \overline{\omega}^2 & \overline{\omega t} - \overline{\omega}\,\overline{t} \\ \overline{\omega t} - \overline{\omega}\,\overline{t} & \overline{t^2} - \overline{t}^2 \end{bmatrix}. \tag{6}$$

It can be seen from the above equation that the information matrix is related to the mean square bandwidth and mean square time width of the transmitted signal. At the same time, according to the parameter estimation theory, the Hessian matrix of the fuzzy function at $\tau = 0, f_d = 0$ is consistent with the Fisher matrix, and the Hessian matrix is obtained as

$$\mathbf{J} = \eta \begin{bmatrix} -\dfrac{\partial^2 \chi(\tau, f_d)}{\partial \tau^2}\Big|_{\substack{\tau = 0 \\ f_d = 0}} & -\dfrac{\partial^2 \chi(\tau, f_d)}{\partial \tau\, \partial f_d}\Big|_{\substack{\tau = 0 \\ f_d = 0}} \\ -\dfrac{\partial^2 \chi(\tau, f_d)}{\partial \tau\, \partial f_d}\Big|_{\substack{\tau = 0 \\ f_d = 0}} & -\dfrac{\partial^2 \chi(\tau, f_d)}{\partial f_d^2}\Big|_{\substack{\tau = 0 \\ f_d = 0}} \end{bmatrix}. \tag{7}$$

Through (7), the process of how to use the waveform parameters and fuzzy functions to solve the Fisher matrix is obtained. During the working process, the music recognition system can perceive the distance and speed information of the target to be observed in real time and obtain the target distance and speed matrix$z = [rv]^T$. This matrix can correspond to the time delay-Doppler frequency estimation matrix at the time of calculation, where $T = \text{diag}[c/2, c/2f_c]$ is the Jacobian matrix of the transformation of the two parameters. Therefore, the measurement noise covariance of distance and speed can be expressed as

$$R = TJ^{-1}T^T. \tag{8}$$

When the parameters of the transmitted signal are determined, the calculation result of (8) can be used as the covariance matrix of the measurement noise and applied to the processing of filtering estimation.

There are many kinds of transmitting signals of the music recognition system. In the simulation of this paper, the transmitting signal of music recognition system in the simulation process is set to be the Gaussian envelope linear frequency modulation signal. The expression for the Gaussian envelope chirp signal is

$$\tilde{s}(t) = \left(\frac{1}{\pi T^2}\right)^{(1/4)} \exp\left(-\left(\frac{-t^2}{2T^2} - jb\right)t^2\right), \tag{9}$$

where b is the frequency modulation slope, and its CRLB is

$$\mathbf{R}(\theta) = \begin{bmatrix} \dfrac{c^2T^2}{2\text{SNR}} & -\dfrac{c^2bT^2}{\omega_c\text{SNR}} \\ -\dfrac{c^2bT^2}{\omega_c\text{SNR}} & \dfrac{c^2}{\omega_c\text{SNR}}\left(\dfrac{1}{2T^2} + 2b^2T^2\right) \end{bmatrix}. \tag{10}$$

Among them, $\theta = [T, b]^T$ is the pulse width and FM slope vector.

The signal pulse width has a direct impact on the signal-to-noise ratio (SNR), and (10) shows that reducing the pulse width can effectively improve the accuracy of distance measurement, and increasing the pulse width can effectively improve the accuracy of speed measurement. If it can analyze the echoes received at the current moment during the working process of the music recognition system and guide the music recognition system to adjust the parameters of the transmitted signal, the music recognition system can change the parameters of the transmitted signal. Moreover, it directly affects the tracking effect of the music recognition system on the target at the next moment, which is the concept and idea of the optimal waveform selection technology. The overall implementation process is shown in Figure 1.

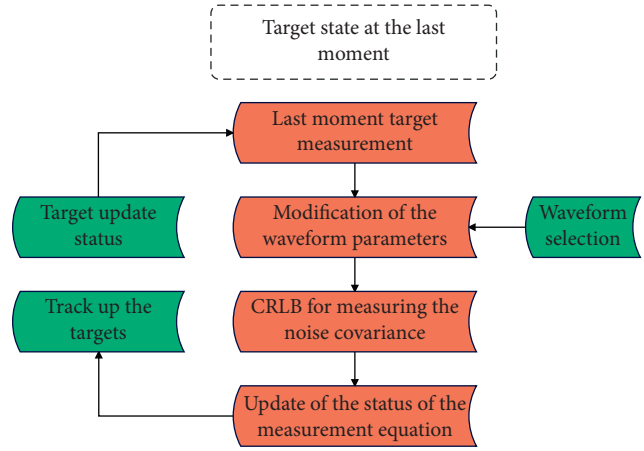


Figure 1: Workflow of the optimal waveform selection technology to achieve target tracking.

$$F_{\text{CV}} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$G_{\text{CV}} = \begin{bmatrix} \dfrac{T^2}{2} & 0 \\ T & 0 \\ 0 & \dfrac{T^2}{2} \\ 0 & T \end{bmatrix}. \tag{12}$$

2.3. Singing Sports Target Model. In this paper, setting the motion state of the singing target includes the following two types: the CT model and the CV model. In the establishment of CT and CV motion models, x and y represent the position of the target in space, and x and y represent the speed of the target in the x and y directions. ω is the turning rate of the target, and T is the sampling period of the music recognition system. The CV model and CT model are as follows.

2.3.1. CV Model. The state variable of the moving target under the CV model is $X = [x, \dot{x}, y, \dot{y}]^{'}$. The state equation of the CV model is

$$\mathbf{X}(k) = \mathbf{F_{CV}}(k-1)\mathbf{X}(k-1) + \mathbf{G_{CV}}(k-1)\mathbf{W}(k-1), \tag{11}$$

where F_{CV} and G_{CV} are, respectively,

$\mathbf{W}(k-1)$ is white Gaussian noise with zero mean.

2.3.2. CT Model. The state variable of the moving target under the CT model is $\mathbf{X} = [x, \dot{x}, y, \dot{y}, \omega]\prime$, and the state equation of the CT model is

$$\mathbf{X}(k) = \mathbf{F}_{\mathbf{CT}}(k-1)\mathbf{X}(k-1) + \mathbf{G}_{\mathbf{CT}}(k-1)\mathbf{W}(k-1), \quad (13)$$

where F_{CT} and G_{CT} are, respectively,

$$F_{\mathrm{CV}} = \begin{bmatrix} 1 & \frac{\sin(\omega T)}{\omega} & 0 & \frac{(\cos(\omega T)-1)}{\omega} & 0 \\ 0 & \cos(\omega T) & 0 & -\sin(\omega T) & 0 \\ 0 & \frac{(1-\cos(\omega T))}{\omega} & 1 & \frac{\sin(\omega T)}{\omega} & 0 \\ 0 & \sin(\omega T) & 0 & \cos(\omega T) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (14)$$

$$G_{\mathrm{CV}} = \begin{bmatrix} \frac{T^2}{2} & 0 & 1 \\ T & 0 & 1 \\ 0 & \frac{T^2}{2} & 1 \\ 0 & T & 1 \end{bmatrix}.$$

$\mathbf{W}(k-1)$ is white Gaussian noise with zero mean.

2.4. IMM Optimal Waveform Selection Algorithm. The traditional IMM optimal waveform selection algorithm includes two parts: the IMM target tracking algorithm and the optimal waveform selection strategy. The main idea of the IMM target tracking algorithm is to use multiple Kalman filters to process the received information at the same time, and each Kalman filter is in a parallel relationship. Therefore, before introducing the IMM algorithm, the Kalman filter algorithm needs to be introduced.

When tracking a moving target, the state space of the target includes two parts: the system equation and the measurement equation. The system equation, expressed in the form of nonlinear equation, describes the actual state of the moving target:

$$x_k = f(x_{k-1}) + v_k, \quad (15)$$

where x_k represents the state of the target at time k, $f(\cdot)$ is the target state transition function, v_k represents the Gaussian white noise caused by the current environment, and the covariance matrix is Q.

The measurement equation is

$$z_k = h(z_{k-1}) + w_k, \quad (16)$$

where z_k represents the tracking vector at time k when the target is being tracked, $h(\cdot)$ represents the target measurement transfer function, w_k represents the Gaussian white noise determined by the waveform parameters, and the covariance matrix is R.

For the Kalman filtering algorithm, when the filtering environment is Gaussian, the Kalman filtering process is consistent with the optimal Bayesian filtering process. Therefore, in the linear environment, the state space equation can be expressed as

$$\begin{aligned} \mathbf{x}_k &= \mathbf{F}\mathbf{x}_{k-1} + \Gamma\mathbf{v}_k, \\ \mathbf{z}_k &= \mathbf{H}\mathbf{x}_k + \mathbf{w}_k. \end{aligned} \quad (17)$$

The flow of the Kalman filter algorithm is as follows. First, the one-step prediction for the target state is

$$\widehat{x}_{k|k-1} = \mathbf{F}\mathbf{x}_{k-1k-1}. \quad (18)$$

The one-step prediction covariance is as follows:

$$\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{\mathbf{k-1|k-1}}\mathbf{F}^T + \Gamma\mathbf{Q}\Gamma^T. \quad (19)$$

The one-step prediction of the measurement is as follows:

$$\widehat{z}_{k|k-1} = \mathbf{H}\widehat{x}_{k|k-1}. \quad (20)$$

The innovation covariance is as follows:

$$\mathbf{S} = \mathbf{H}_{k|k-1}\mathbf{H}^T + \mathbf{R}. \quad (21)$$

The Kalman filter gain is as follows:

$$\mathbf{K} = \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}^{-1}. \quad (22)$$

When the time is k, the state equation of the target can be updated as follows:

$$\widehat{x}_{k|k} = \widehat{x}_{k|k-1} + \mathbf{K}\left(z - \mathbf{H}\widehat{x}_{k|k-1}\right). \quad (23)$$

When the time is k, the covariance matrix of the filtering error is as follows:

$$\mathbf{P}_{k|k} = [\mathbf{I} - \mathbf{KH}]\mathbf{P}_{k|k-1}[\mathbf{I} - \mathbf{KH}]^T + \mathbf{KRK}^T. \quad (24)$$

As an effective technique for maneuvering target tracking, the IMM algorithm uses multiple model sets to describe the possible model states of the system. The basic idea of this algorithm is that, when each model is valid at the current moment at different times, the initial conditions of the filter matching the model are obtained by comparing the estimated values of the states obtained by all filters at the previous moment. The basic filtering steps are implemented in parallel for each model. Finally, the model matching likelihood function is used as the basis to update the model probability, and the estimated value of the state is obtained by the weighted summation of all the modified state estimates of the filters. The algorithm obtains the final tracking result by mixing the estimated values obtained from different models.

The IMM algorithm mainly consists of the following four steps: input interaction, filtering, model probability update,

and output synthesis. The specific process of each step is as follows:

2.4.1. Input Interaction. The input interaction is obtained by the initial value of the filter period of each filter from the model conditional transition probability and state estimation value, and the transition probability between models is set as $P_{t_{ij}}$.

$$\begin{aligned} P^{oj}(k-1|k-1) &= \sum_{i=1}^{r} \mu_{ij}(k-1|k-1) \\ &\cdot \left\{ P^{i}(k-1|k-1) + \left[\widehat{X}^{i}(k-1|k-1) - \widehat{X}^{oj}(k-1|k-1) \right] \bullet \left[\widehat{X}^{i}(k-1|k-1) - \widehat{X}^{oj}(k-1|k-1) \right]^{T} \right\}, \\ \mu_{ij}(k-1|k-1) &= P\left\{ \frac{M_i(k-1)}{M_i(k), Z^{k-1}} \right\}, \\ &= \frac{p_{ij}\mu_i(k-1)}{\overline{c}_j}, \end{aligned} \tag{26}$$

where $j = 1, \ldots, r$, p_{ij} is the transition probability from model i to model j, and $\overline{c}_j$ is the normalization constant, $\overline{c}_j = \sum_{i=1}^{r} p_{ij}\mu_i(k-1)$.

2.4.2. Model Filtering. According to the calculated model input state and error covariance, combined with the observation data at the current moment, the Kalman filter is performed to obtain the filter output of each model.

2.4.3. Update of Model Probability.

$$\begin{aligned} \mu_j(k) &= P\left\{ \frac{M_j(k)}{Z^k} \right\}, \\ &= P\left\{ \frac{Z(k)}{M_j(k)}, Z^{k-1} \right\} P\left\{ \frac{M_j(k)}{Z^{k-1}} \right\}, \\ &= \frac{1}{c}\Lambda_j(k) \sum_{i=1}^{r} p_{ij}\mu_i(k-1), \\ &= \frac{\Lambda_j(k)\overline{c}_j}{c}, \end{aligned} \tag{27}$$

where c is a normalization constant, $c = \sum_{j=1}^{r} \Lambda_j(k)\overline{c}_j$ and $\Lambda_j(k)$ is the likelihood function of observation Z(k), where $\Lambda_j(k)$ represents the possibility of model j, and the expression of $\Lambda_j(k)$ is

The input state of the resulting computational model is as follows:

$$\widehat{X}^{oj}(k-1|k-1) = \sum_{i=1}^{r} \widehat{X}^{i}(k-1|k-1)\mu_{ij}(k-1|k-1). \tag{25}$$

The error covariance is as follows:

$$\begin{aligned} \Lambda_j(k) &= P\left\{ \frac{Z(k)}{M_j(k)}, Z^{k-1} \right\}, \\ &= \frac{1}{(2\pi)^{n/2}\left|\mathbf{S}_j(k)\right|^{(1/2)}} \exp\left\{ -\frac{1}{2}\mathbf{v}_j^{T}\mathbf{s}_j^{-1}(k)\mathbf{v}_j \right\}, \end{aligned} \tag{28}$$

where v_j is the filter residual estimation of model j, $\mathbf{S}_j$ is the covariance matrix, which obeys the Gaussian distribution, and the expression is

$$\begin{aligned} \mathbf{v}_j(k) &= \mathbf{Z}(k) - \mathbf{H}(k)\widehat{X}^{j}(k|k-1), \\ \mathbf{S}_j(k) &= \mathbf{H}(k)\mathbf{P}_{\overline{X}}^{j}(k|k-1)\mathbf{H}^{T}(k) + \mathbf{R}(k). \end{aligned} \tag{29}$$

2.4.4. Output Interaction. All model states are weighted, and the output of the system state estimation at time k is obtained by using the product of the model probability of each model and the state estimate value:

$$\begin{aligned} \widehat{\mathbf{X}}\left(\frac{k}{k}\right) &= \sum_{j=1}^{r} \widehat{X}^{j}(k|k)\mu_j(k), \\ \mathbf{P}\left(\frac{k}{k}\right) &= \sum_{j=1}^{r} \mu_j(k) \\ &\cdot \left\{ \mathbf{P}^{j}\left(\frac{k}{k}\right) + \left[\widehat{X}^{j}\left(\frac{k}{k}\right) - \widehat{X}\left(\frac{k}{k}\right) \right] \left[\widehat{X}^{j}\left(\frac{k}{k}\right) - \widehat{X}\left(\frac{k}{k}\right) \right]^{T} \right\}. \end{aligned} \tag{30}$$

Through the calculation process of the IMM algorithm, it can be seen that the IMM algorithm achieves the best global tracking performance by calculating the weighted sum value

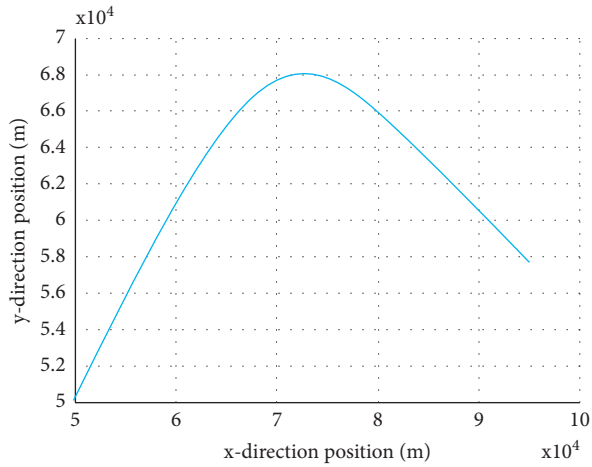


Figure 2: Target movement trajectory.

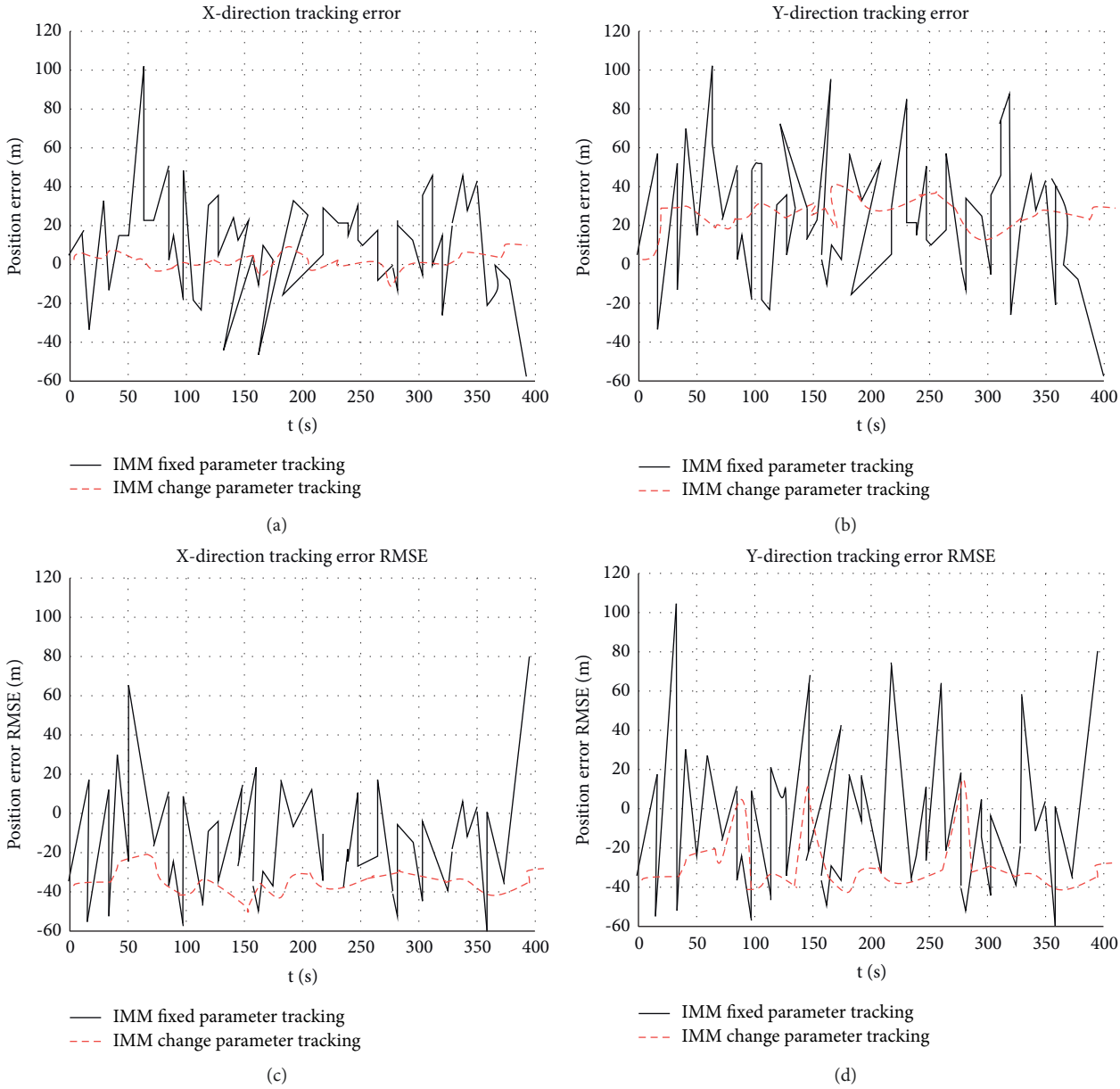
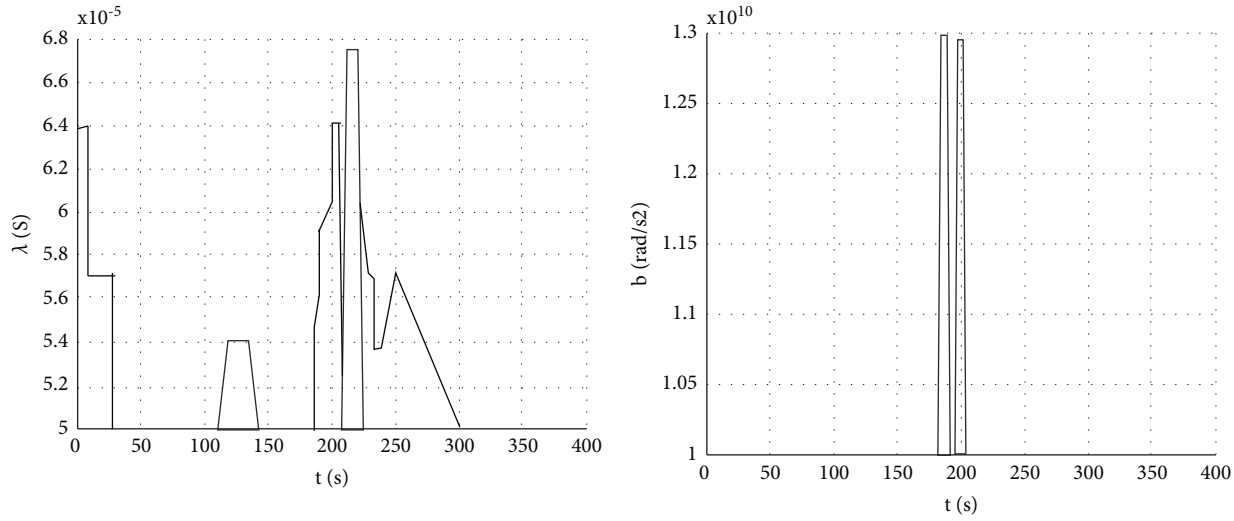


Figure 3: Comparison of IMM optimal waveform selection algorithm and IMM algorithm on target tracking effect when the waveform is fixed: (a) x-direction tracking distance error, (b) y-direction tracking distance error, (c) x-direction tracking distance mean error, and (d) y-direction tracking distance mean error.



(a) (b)

Figure 4: Changes of waveform parameters transmitted by IMM optimal waveform selection algorithm. (a) Change diagram of the pulse width during the tracking process. (b) Changes of FM parameters during tracking.

of each model tracking estimated value and the model matching update probability during the entire tracking process.

The traditional selection of the optimal waveform using the IMM optimal waveform selection algorithm is based on the filtering output result of the IMM target tracking algorithm. Moreover, it sets a certain waveform selection criterion function and determines the next moment waveform according to the waveform selection criterion function. In the tracking task, the more commonly used waveform selection criterion function is the mean square minimum error criterion; that is, in each moment of tracking, the mean square value of the output state estimation error needs to be minimized. Therefore, the waveform selection criteria used in the IMM optimal waveform selection algorithm are as follows:

$$\theta_k^* = \text{argmin}_{\theta_k \in \theta} \text{Tr}\{\mathbf{P}_{k/k}(\theta_k)\}, \tag{31}$$

where θ_k^* represents the waveform selected at time k and the filtering covariance at time k needs to be minimized, namely,

$$\begin{aligned} &\text{Tr}\{\mathbf{P}_{k/k}(\theta_k)\}, \\ &= \text{Tr}\left\{\mathbf{P}_{k/k} - \mathbf{P}_{k|k-1}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}(\theta_k)\right)^{-1}\mathbf{H}\mathbf{P}_{k|k-1}\right\}. \end{aligned} \tag{32}$$

From the above equation, it can be known that at time k, the trace of the filter covariance is a variable related to the measurement noise covariance. According to (31) and (32), the parameters of the music recognition system at the next moment can be obtained.

3. Simulation Experiments and Experimental Results

3.1. Target Parameters and Music Recognition System Parameters. In this section, the transmitter waveform used in the simulation is a Gaussian envelope chirp waveform with a carrier frequency of 2 GHz. Because the parameters of the music recognition system are variable, the waveform library of the music recognition system composed of all the waveforms of different parameters is

$$\mathfrak{R} = \{\lambda \in [50e-6: 3.5e-6: 120e-5], b \in [1e10: 1e9: 1el1]\}. \tag{33}$$

From the composition formula of the waveform library, it can be calculated that the waveform library used in the simulation is composed of 651 waveforms. At the same time, in order to compare the effect, a music recognition system with fixed parameters is set up. The parameters of the music recognition system are as follows: the frequency modulation parameter is $b = 1 \times 10^{10}\text{rad/s}^2$, and the pulse width is $\lambda = 5 \times 10^{-5}\text{s}$. The tracked target does a uniform linear motion (CV model) within 1150 seconds and a uniform curve motion (CT model) for 151270 seconds. In the 271400th second, the target's motion trajectory is still a uniform linear motion (CV model). Taking the music recognition system as the coordinate origin to establish a coordinate system, the position coordinate (x_0, y_0) of the initial movement of the target is (50000, 50000), and the initial speed (v_{x0}, v_{y0}) is (100, 100). The angular velocity of the set target when moving with the CT model is $\omega = -(\pi/270)$. The trajectory of the target is shown in Figure 2.

3.2. IMM Algorithm Parameters. In the filtering of the IMM algorithm, two model sets, the CT model and CV model are used, and the transition matrix between the two models is set as $\mathbf{P} = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$, and 50 Monte Carlo simulations are

Table 1: Accuracy of music skill recognition.

Number	Singing recognition (%)
1	85.69
2	80.18
3	88.35
4	77.51
5	84.43
6	77.51
7	80.60
8	86.62
9	88.66
10	80.76
11	84.56
12	82.28
13	81.26
14	86.77
15	80.51
16	78.42
17	75.71
18	81.92
19	77.77
20	79.80
21	87.84
22	88.66
23	82.80
24	87.23
25	74.85
26	84.14
27	88.47
28	76.83
29	83.47
30	82.73
31	76.11
32	81.33
33	86.49
34	87.31
35	82.92
36	86.42
37	76.99
38	86.19
39	79.47
40	75.71
41	88.96
42	83.07
43	78.31
44	78.29
45	76.52
46	74.78
47	80.53
48	84.98
49	88.53
50	82.85
51	74.52
52	86.99
53	81.88
54	77.17
55	86.65
56	86.08
57	78.19
58	80.39
59	86.10
60	86.37

Table 2: The improvement effect of music teaching.

Number	Teaching effect
1	63.35
2	61.53
3	68.68
4	71.91
5	72.52
6	71.41
7	69.82
8	58.39
9	57.09
10	55.60
11	60.48
12	69.49
13	55.81
14	73.54
15	61.10
16	67.93
17	64.69
18	62.70
19	59.64
20	63.69
21	74.64
22	57.21
23	75.98
24	68.88
25	71.29
26	74.94
27	60.40
28	58.76
29	58.37
30	75.33
31	74.35
32	56.11
33	65.95
34	58.82
35	67.25
36	65.72
37	55.60
38	62.50
39	75.47
40	66.81
41	63.35
42	72.18
43	55.42
44	72.25
45	67.78
46	74.13
47	63.32
48	67.73
49	74.84
50	64.67
51	73.58
52	75.21
53	73.03
54	68.98
55	62.15
56	71.06
57	56.98
58	56.04
59	69.90
60	67.72

carried out. The tracking results obtained by the IMM algorithm under the transmission waveform parameters and waveform selection of the fixed music recognition system, as well as the speed and distance errors obtained in the x and y directions, respectively, are obtained, as shown in Figure 3.

It can be seen from Figure 3 that the waveform selection can be realized by using the IMM optimal waveform selection algorithm, and the tracking error can be reduced when the waveform parameters change in real time compared to the fixed waveform. Figure 4 shows the changes in the parameters of the waveform transmitted by the music recognition system during the tracking process.

It can be seen from Figure 4 that the algorithm only utilizes a small number of waveform parameters, which wastes a lot of waveform resources, which will affect the effect of target tracking. Therefore, we consider adjusting the waveform selection criterion function to increase the efficiency of using the waveform and improve the tracking effect of the target.

On the basis of the above simulation experiments, the method based on music waveform feature extraction proposed in this paper is applied to the practical application of music education singing skills. The music singing skills are verified, and the recognition accuracy of music skills and the improvement effect of music teaching are counted, and the results shown in Tables 1 and 2 are obtained.

From the above research results, it can be seen that the intelligent analysis system of music education singing skills based on music waveform feature extraction proposed in this paper has good effects and can effectively promote the improvement of music singing skills and improve the effect of music teaching.

4. Conclusion

The visualization of music information is relevant to many disciplines. It takes music and acoustics as its scientific basis and draws scientific nutrition from basic disciplines such as physics, mathematics, and psychology. Moreover, it is supported by engineering technologies such as human-computer interaction, sensing, and control technology, computer and information technology. In addition, it is closely related to music iconography and imaging, and its final application fields include music creation, music performance, music research, music education, etc. The development of music information visualization is a dynamic process, which must be studied and analyzed in combination with the social and cultural background and the level of scientific and technological development in each historical period. In this paper, the intelligent analysis of singing skills in music education is carried out with the support of music waveform feature extraction technology. From the research results, it can be seen that the intelligent analysis system of music education singing skills based on the extraction of music waveform features has good effects and can effectively promote the improvement of music singing skills and the effect of music teaching.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This study was sponsored by the Xinyang Vocational and Technical College.

References

[1] F. Calegario, M. M. Wanderley, S. Huot, G. Cabral, and G. Ramalho, "A method and toolkit for digital musical instruments: generating ideas and prototypes," *IEEE MultiMedia*, vol. 24, no. 1, pp. 63–71, 2017.

[2] D. Tomašević, S. Wells, I. Y. Ren, A. Volk, and M. Pesek, "Exploring annotations for musical pattern discovery gathered with digital annotation tools," *Journal of Mathematics and Music*, vol. 15, no. 2, pp. 194–207, 2021.

[3] X. Serra, "The computational study of a musical culture through its digital traces," *Acta Musicologica*, vol. 89, no. 1, pp. 24–44, 2017.

[4] I. B. Gorbunova and N. N. Petrova, "Digital sets of instruments in the system of contemporary artistic education in music: socio-cultural aspect," *Journal of Critical Reviews*, vol. 7, no. 19, pp. 982–989, 2020.

[5] E. Partesotti, A. Peñalba, and J. Manzolli, "Digital instruments and their uses in music therapy," *Nordic Journal of Music Therapy*, vol. 27, no. 5, pp. 399–418, 2018.

[6] B. Babich, "Musical "covers" and the culture industry," *Research in Phenomenology*, vol. 48, no. 3, pp. 385–407, 2018.

[7] L. L. Gonçalves and F. L. Schiavoni, "Creating digital musical instruments with libmosaic-sound and mosaicode," *Revista de Informática Teórica e Aplicada*, vol. 27, no. 4, pp. 95–107, 2020.

[8] I. B. Gorbunova, "Music computer technologies in the perspective of digital humanities, arts, and researches," *Opción*, vol. 35, no. SpecialEdition24, pp. 360–375, 2019.

[9] A. Dickens, C. Greenhalgh, and B. Koleva, "Facilitating accessibility in performance: participatory design for digital musical instruments," *Journal of the Audio Engineering Society*, vol. 66, no. 4, pp. 211–219, 2018.

[10] O. Y. Vereshchahina-Biliavska, O. V. Cherkashyna, Y. O. Moskvichova, O. M. Yakymchuk, and O. V. Lys, "Anthropological view on the history of musical art," *Linguistics and Culture Review*, vol. 5, no. S2, pp. 108–120, 2021.

[11] A. C. Tabuena, "Chord-interval, direct-familiarization, musical instrument digital interface, circle of fifths, and functions as basic piano accompaniment transposition techniques," *International Journal of Research Publications*, vol. 66, no. 1, pp. 1–11, 2020.

[12] L. Turchet and M. Barthet, "An ubiquitous smart guitar system for collaborative musical practice," *Journal of New Music Research*, vol. 48, no. 4, pp. 352–365, 2019.

[13] R. Khulusi, J. Kusnick, C. Meinecke, C. Gillmann, J. Focht, and S. Jänicke, "A survey on visualizations for musical

data," *Computer Graphics Forum*, vol. 39, no. 6, pp. 82–110, 2020.

[14] E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley, and F.-R. Stoter, "Musical source separation: an introduction," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 31–40, 2019.

[15] T. Magnusson, "The migration of musical instruments: on the socio-technological conditions of musical evolution," *Journal of New Music Research*, vol. 50, no. 2, pp. 175–183, 2021.

[16] I. B. Gorbunova and N. N. Petrova, "Music computer technologies, supply chain strategy and transformation processes in socio-cultural paradigm of performing art: using digital button accordion," *International Journal of Supply Chain Management*, vol. 8, no. 6, pp. 436–445, 2019.

[17] J. A. A. Amarillas, "Marketing musical: música, industria y promoción en la era digital," *INTERdisciplina*, vol. 9, no. 25, pp. 333–335, 2021.

[18] G. Scavone and J. O. Smith, "A landmark article on nonlinear time-domain modeling in musical acoustics," *Journal of the Acoustical Society of America*, vol. 150, no. 2, pp. R3–R4, 2021.