Hindawi
Advances in Mathematical Physics
Volume 2021, Article ID 8678853, 12 pages
https://doi.org/10.1155/2021/8678853

Research Article

Feature Extraction of Music Signal Based on Adaptive Wave Equation Inversion

Tianzhuo Gong [1,2] and Sibing Sun [2]

[1] *Music Collage, Capital Normal University, Beijing 100000, China*
[2] *School of Music, Harbin Normal University, Harbin 150080, China*

Correspondence should be addressed to Sibing Sun; deanmusic@hrbnu.edu.cn

Received 13 September 2021; Accepted 9 October 2021; Published 22 October 2021

Academic Editor: Miaochao Chen

The digitization, analysis, and processing technology of music signals are the core of digital music technology. There is generally a preprocessing process before the music signal processing. The preprocessing process usually includes antialiasing filtering, digitization, preemphasis, windowing, and framing. Songs in the popular wav format and MP3 format on the Internet are all songs that have been processed by digital technology and do not need to be digitalized. Preprocessing can affect the effectiveness and reliability of the feature parameter extraction of music signals. Since the music signal is a kind of voice signal, the processing of the voice is also applicable to the music signal. In the study of adaptive wave equation inversion, the traditional full-wave equation inversion uses the minimum mean square error between real data and simulated data as the objective function. The gradient direction is determined by the cross-correlation of the back propagation residual wave field and the forward simulation wave field with respect to the second derivative of time. When there is a big gap between the initial model and the formal model, the phenomenon of cycle jumping will inevitably appear. In this paper, adaptive wave equation inversion is used. This method adopts the idea of penalty function and introduces the Wiener filter to establish a dual objective function for the phase difference that appears in the inversion. This article discusses the calculation formulas of the accompanying source, gradient, and iteration step length and uses the conjugate gradient method to iteratively reduce the phase difference. In the test function group and the recorded music signal library, a large number of simulation experiments and comparative analysis of the music signal recognition experiment were performed on the extracted features, which verified the time-frequency analysis performance of the wave equation inversion and the improvement of the decomposition algorithm. The features extracted by the wave equation inversion have a higher recognition rate than the features extracted based on the standard decomposition algorithm, which verifies that the wave equation inversion has a better decomposition ability.

1. Introduction

Music can express people's thoughts and can convey people's happiness, anger, sorrow, and joy. It exists in various cultures and countries and is closely related to people's lives [1]. Since the reform and opening up, music has been constantly developing and changing, and there have been many different styles of music and a large number of music works [2]. 64% of users cannot find the song they want to listen to when using a music search engine, and many users are not clear about their needs for music [3]. In the face of the large number of music, it is difficult to find your favorite music, and music classification and search still have huge room for development [4]. Today, with the popularization of the Internet and the continuous development of network applications, in the face of a huge user group and massive scale of data, the importance of digital music retrieval and recommendation is self-evident [5]. Music classification is an important field of music retrieval. It is the premise, technical means, and main work content of the research on content-based music retrieval and recommendation. The study of music style classification has a broad development space [6]. The classification of music styles can help people quickly find their favorite music and can play different styles of music at different times according to different occasions.

In music signal recognition technology, the key issue is to establish an acoustic model of music signal recognition primitives [7]. At present, some technologies for the acoustic model of music signals have not been completely solved, resulting in the performance of some products that are still difficult to meet the ideal use requirements [8]. The acoustic model is established on the basis of the characteristic parameters of the music signal. Therefore, the amount of useful information contained in the characteristic parameters of the music signal directly determines the accuracy of the acoustic model's description of the music signal. The parameter information is less, so the final acoustic model is also imperfect. Before the acoustic model is established, the characteristic parameters must be studied to extract the parameters with the most useful information [9]. Acoustic models established based on characteristic parameters mainly fall into two categories. One is mapping planning on the time axis, and the distortion of the two characteristic parameters is measured; the other is a model based on statistical knowledge. The establishment is based on the initial model and training data and constantly reevaluates and optimizes the parameters until the model converges. This algorithm is not a global optimal analytical solution, and it is easier to fall into a local optimal solution. The final model parameters are quite different. Therefore, the study of feature parameter extraction and model initialization is of great significance in music signal recognition.

This article introduces the adaptive wave equation inversion method as the core and analyzes the wave equation inversion in the time domain, including the idea of wave equation inversion. The objective function is given for the full-wave equation inversion method in the time domain, and the local optimization algorithm, namely the gradient method, is used for inversion. The gradient formula is given, and a detailed derivation process is attached. Aiming at the period jump, the proposed adaptive wave equation inversion is introduced, including the basic principle and objective function of the method. A new objective function is used to give the formulas of the accompanying source and gradient and give a detailed derivation process. The gradient difference between adaptive wave equation inversion and full-wave equation inversion is compared. The gradient formula of the conjugate gradient method and the selection of the step length are introduced. We perform vowel recognition experiments on the music signal libraries 1 and 2, respectively. For features of the same dimension, on music signal library 1, the wave equation inversion has a higher recognition rate than the three contrasted features. On music signal library 2, it also has a higher recognition rate under the original signal and low signal-to-noise ratio. For the combination of features in this article, HMS-MFCC has a strong characterization ability, while EWCF is more susceptible to noise pollution, but it has the lowest dimensionality.

2. Related Work

The purpose of feature extraction is to obtain information that is conducive to identification and eliminate interference in the music signal. The music signal contains a large amount of not only music signal information but also personal characteristic information. The characteristic parameters of the music signal should be able to accurately represent all the information contained in the original signal that helps to distinguish. Thorough research makes the existing characteristic parameters unable to completely and accurately characterize the information of the music signal. At present, the characteristic parameters in music signals can be divided into time domain, frequency domain, and cepstrum domain. The time domain parameters are obtained by reducing the dimensionality of each frame of music signal in the time domain to form a set of feature vectors. Time domain parameters mainly include short-term energy, short-term zero-crossing rate, and autocorrelation coefficient. The frequency domain and cepstrum domain parameters are to transform each frame of music signal into the frequency domain range and extract characteristic parameters in the frequency domain or convert the frequency domain parameters into the cepstrum domain.

At present, there is no parameter in feature extraction that can represent all useful information of music signals, even if the more mature MFCC parameters are used [10]. Among the various parameters, it is an approximate description of a certain aspect of the music signal. For example, the commonly used MFCC parameters simulate the human auditory system, which mainly considers low-frequency components. The low-frequency components of the parameters account for the main part, and the use of the differences of the components of the MFCC parameters is not considered for feature selection, so that the parameters will lose some important information [11]. Researchers have proposed many algorithms to improve the characteristic parameters of music signals [12].

The Mel-frequency cepstral coefficient is currently the most widely used characteristic coefficient in the music signal recognition system. It is based on the auditory system of the human ear and extracts parameters by simulating the auditory system of the human ear to establish a model to describe the energy distribution of the music signal in the frequency domain [13]. For sounds of different frequencies, the ability of the human auditory system to perceive them is different. For sounds with a frequency below 1,000 Hz, the auditory system's ability to perceive it satisfies an approximate linear relationship, but when the frequency is higher than 1,000 Hz, the auditory system's perception of sound meets a logarithmic relationship with the frequency approximately [14]. Compared with PLC and PLCC parameters, MFCC parameters emphasize the low-frequency information of music signals, shielding high-frequency noise interference, and without any assumptions can be used in various situations.

With the advancement and development of computer science and technology, the basic theories and key technologies of music signal recognition technology have been initially promoted [15]. The main research results of music signal recognition technology during this period are dynamic programming (DP) and linear prediction (LP). Among them, the dynamic programming technology is a technology to calibrate a group of music signals in time. It can better solve the problem of unequal length correction of music signals in music signal recognition [16]. The linear predictive analysis technology proposes a better solution to the mathematical model of music signal generation, which has a profound impact on the development and application of music signal recognition technology [17]. At the same time, NEC Laboratory, Tokyo Radio Laboratory in Japan, and Kyoto University have successively researched and produced dedicated hardware devices to be used in music signal recognition technology, laying a solid foundation for their further theoretical research and practical application [18, 19].

The Baum-Welch algorithm is essentially an algorithm that uses the maximum expected value [20]. This algorithm can ensure that the output probability of the model that is not reevaluated once is increased, but this algorithm has a large dependence on the initial parameters. For different initial parameters, the final output probability is not unique. Therefore, the traditional Baum-Welch algorithm cannot accurately and completely establish an acoustic model of the trained music signal observation sequence [21]. In terms of the hidden Markov models, how to train a perfect acoustic model has always been a difficult point in research [22]. In order to solve the problem that the Baum-Welch algorithm's dependence on the initial model parameters may cause the final training model to fall into a local optimum, researchers have proposed various solutions and algorithms [23]. These algorithms are mainly aimed at two aspects: one is in the algorithm training process, combined with other algorithms, to intelligently optimize the model parameters obtained from each revaluation. These algorithms generally have the advantages of global optimization [24]. The other is to optimize parameters in the model initialization stage and try to choose more appropriate model initialization parameters [25].

3. Music Signal Processing Technology

3.1. Mathematical Model of Music Signal. Based on the characteristics of the vocal tract model of the music signal, the music signal model is composed of three parts: (1) glottal excitation function $G(z)$, (2) vocal tract modulation function $V(z)$, and (3) lip radiation function $R(z)$.

The music signal generation system is formed by connecting these three functions in series, namely

$$H(z) = R(z)G(z)V(z). \tag{1}$$

Common vocal tract models include lossless sound tube and formant models. The excitation wave of the sound source is affected by the resonance of the vocal tract, and resonance occurs in certain frequency bands. The peak produced by the envelope of the spectral line at the resonant frequency is the resonant peak. The vocal tract model of general vowels is represented by the all-pole model, and the nongeneral vowels and most consonants are represented by the zero-pole model. The transfer function expression of a second-order resonator is

$$V_i(z) = \frac{A_i}{1 - B_i z^{-1} - C_i z^{-2}}. \tag{2}$$

Multiple V_i linear combinations are obtained to obtain the formant model of the sound channel:

$$V(z) = \lim_{M \longrightarrow \infty} \prod_{i=0}^{M-1} \left[\left(1 - A_i\right) \times \left(1 - B_i z^{-1} + C_i z^{-2}\right) \right]. \tag{3}$$

Since the excitation model of the music signal is an expression in the form of all poles, we call the ratio of the music signal to the output wave velocity of the vocal tract as the radiation impedance, ignoring that the open area of the lips is much smaller than the head surface area, and derive the radiation impedance expression:

$$Z_L(\Omega) = jL_r R_r \Omega \times \left(R_r - jL_r \Omega\right)^{-1}. \tag{4}$$

In the actual process, the physical process of music signal generation is different from the above three models but is approximately equivalent. This also verifies that the music signal is a short-term stable signal and a signal that changes over time. In addition, the fricatives in voiced sounds have both unvoiced and voiced excitation sources at the same time and cannot be obtained by simply superimposing the two.

3.2. Preprocessing of Music Signal. The music signal is represented by a time-varying function curve on the mathematical image, and its dimension is $N \times 1$, which is a column vector. Among them, N is the sum of the number of samples in the music signal. Through sampling and A/D conversion, the music signal is changed from an analog signal to a digital signal. The sampling rate is the number of times the music signal is sampled within 1 s per unit time. The higher the sampling rate, the more music signal information is obtained per unit time. The restoration of the music signal is more real. In order to maintain the maximum characteristics of the music signal and avoid spectrum aliasing, the Nyquist theorem must be satisfied when sampling, that is, the sampling frequency $fs > 2fm$, and fm is the highest frequency of the music signal. Quantization is to divide the amplitude of the entire range into a finite set, specify the waveform of one of the ranges as the standard, and treat the amplitude of all the waveforms as having the same amplitude as that.

The preemphasis processing is to consider that the music signal in the high-frequency band above 800 Hz has a 6 dB/octave amplitude drop. Sometimes, it is also considered to eliminate the DC level offset, so the high-frequency part of the music signal must be added through a transfer function.

The music signal is a short-term stable signal, and its characteristics can be considered to remain unchanged within 10 ms. The part of the sound interval obtained by multiplying the music signal by the window function is called a frame. The length of the interval is called the frame length. Generally, there are 33-100 frames per second. The overlapping part between adjacent frames is called a frame. In order to make a continuous and smooth transition between frames, the frame shift is usually 1/3 of the frame length.

The main lobe of the rectangular window is narrow and sharp, and the corresponding frequency resolution is high, the side lobe peak is large, and the spectral smoothing effect is good, but the spectrum leakage is more serious; the width of the main lobe of the Hamming window is large, which can greatly retain the waveform characteristics of the music signal. But its side lobe attenuation is relatively large. According to the music signal waveform multiplied by the window function, there will be no sharp changes, and the music signal waveform characteristics should be maintained as much as possible. When selecting the window, the main lobe width, frequency resolution, and side lobe attenuation should be comprehensively considered.

Endpoint detection can find out the start and end of the sound segment in the signal, which can remove the silent segment, enhance the useful part of the signal, and reduce the length of the voice. For isolated word recognition, the main purpose is to reduce the amount of calculation and noise interference and increase the calculation accuracy; for continuous speech recognition, it is mainly used to divide the recognition primitives and to model and recognize the recognition primitives. Only by accurately finding the starting end of the voice signal can the subsequent processing of the voice be accurately performed. The schematic diagram of dual-threshold method endpoint detection is shown in Figure 1.

4. Mathematical Model of Adaptive Wave Equation Inversion

4.1. Adaptive Wave Equation Inversion. The objective function of the full-wave equation inversion in the time domain is

$$C(m) = 0.5(\Delta d)^2 = 0.5(d - u)^2. \tag{5}$$

The forward simulation wave field is u, the wave field is d, and δd is the residual of the two. The residual equation of the wave equation inversion is

$$\Delta d_i = \mathrm{Sup}\{u_i - d_i\}. \tag{6}$$

When the phase difference between the predicted data and the real data is greater than half a cycle, a cycle jump will occur at this time. When used in actual seismic data, because the initial model is not so accurate in most cases, it is prone to cycle jumping, which has a great impact on the inversion. Based on this, we proposed to introduce a penalty term to constrain the objective function to overcome the cycle jump.

Figure 2 is a schematic diagram of cycle skip artifacts in FWI. The solid blue line represents the time function of the true waveform of period T. The solid red line above represents the predicted waveform with a time delay greater than $T/2$ cycles from the real waveform. In this case, FWI will update the underground medium model so that the seismogram of the $(n+1)$th period predicted data will match the nth period of the observation data map. An error occurs in the update of the underground medium model, resulting in the inversion effect, deviation. In the example at the bottom, the n periods of the predicted data and the observed data are consistent, because the time delay is less than $T/2$, and the FWI can get the correct underground medium model updated.

The adaptive wave equation inversion is proposed to suppress the influence of cycle jumps on the inversion, and it can be inverted under an unsatisfactory initial model to obtain relatively still ideal inversion results.

The theory and method of adaptive wave equation inversion are different from the traditional full-wave equation inversion method. Here, the convolution of the filter and one of the data sets is used to subtract from the other data set instead of direct subtraction. The adaptive full-wave equation inversion can well suppress the occurrence of cycle jumps.

The convolution of a signal $f(t)$ and the impact signal $\delta(t)$ is equal to $f(t)$ itself. When the wave field value d is convolved with the shock function, the wave field d is obtained. When the predicted wave field data u is very close to the real wave field data, $u \cdot \delta = d$ can be obtained. The filter coefficients are calculated, and the simulated data is convolved with the filter coefficients. Through continuous iteration, the simulated data keeps getting closer to the real data, and at the same time, the phase difference between the two is gradually reduced, and the cycle jump is well suppressed. The filter coefficient gradually becomes a shock function or approximates the shock function. At this time, the difference between the simulated data and the real data is minimized, and finally, an ideal inversion effect is achieved. This method is called forward adaptive fluctuation equation inversion. At the same time, when the real data is convolved with the filter coefficients and then compared with the simulated data, the gap between the two can also be reduced through iteration. This method is called the subsequent adaptive wave equation inversion.

4.2. Inversion of Objective Function by Adaptive Wave Equation. The objective function of adaptive wave equation inversion is different from that of traditional full-wave equation inversion. With dual objective functions, the inversion is also divided into two steps: the first step is to calculate the filter coefficients. The second step is to determine the new accompanying source through the filter coefficients and calculate the gradient combined with the step size for iterative calculation. The first step is to design a Wiener filter here, that is, to define a Wiener filter w of order l, first convolve the filter with the real data, and then the result of the convolution with the least squares of the simulated data

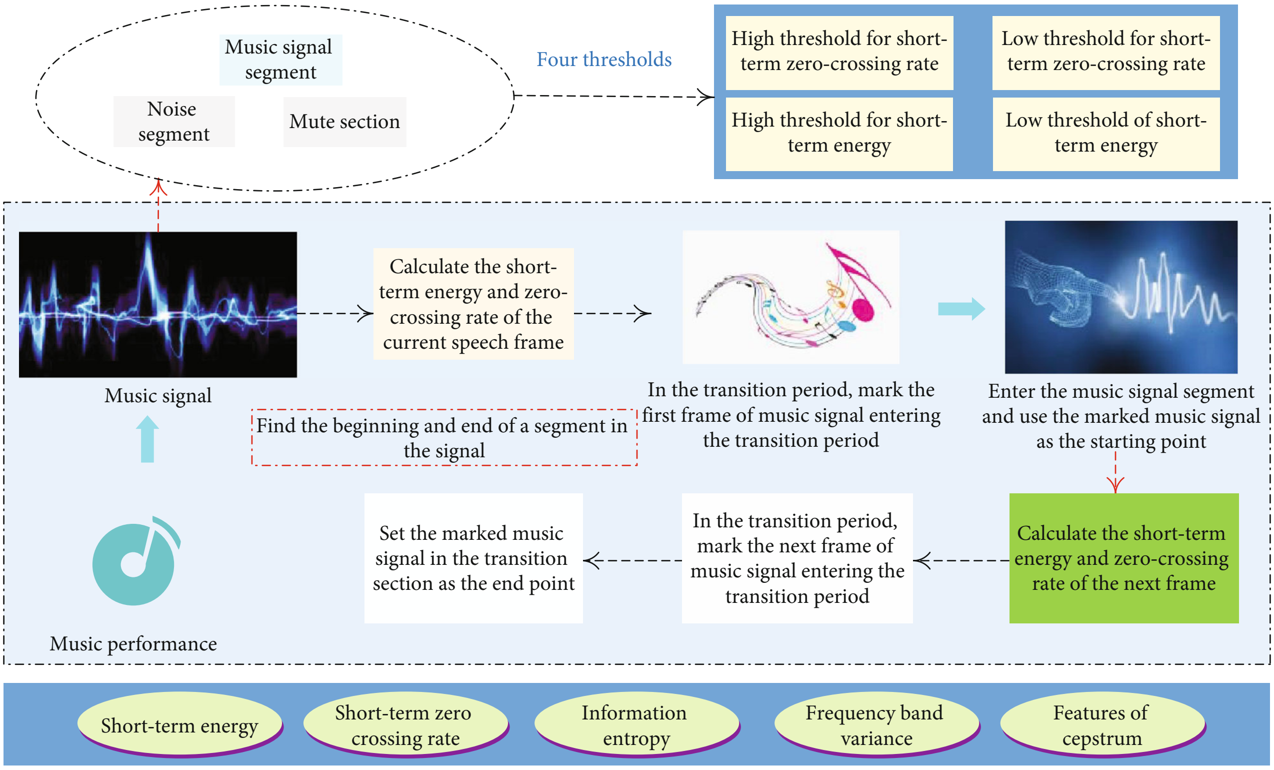
Music signal segment
Noise segment
Mute section
Four thresholds
High threshold for short-term zero-crossing rate
Low threshold for short-term zero-crossing rate
High threshold for short-term energy
Low threshold of short-term energy
Music signal
Calculate the short-term energy and zero-crossing rate of the current speech frame
In the transition period, mark the first frame of music signal entering the transition period
Enter the music signal segment and use the marked music signal as the starting point
Find the beginning and end of a segment in the signal
Music performance
Set the marked music signal in the transition section as the end point
In the transition period, mark the next frame of music signal entering the transition period
Calculate the short-term energy and zero-crossing rate of the next frame
Short-term energy
Short-term zero crossing rate
Information entropy
Frequency band variance
Features of cepstrum

Figure 1: Schematic diagram of dual-threshold method endpoint detection.

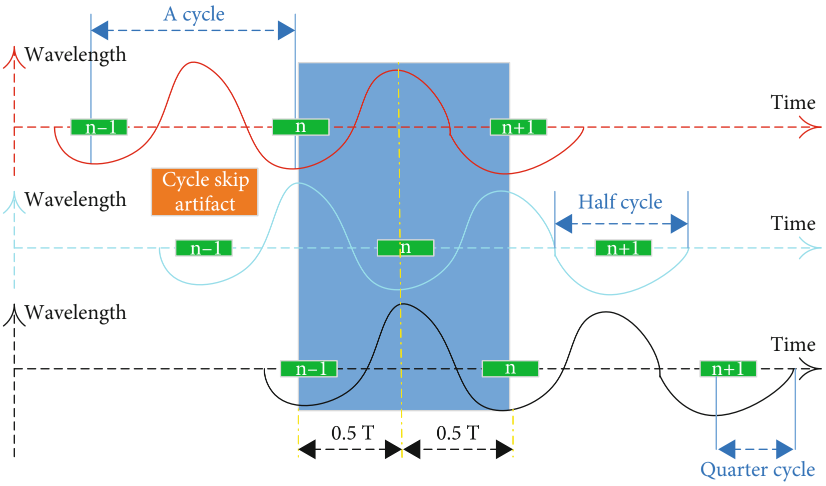

Figure 2: Schematic diagram of cycle jump.

can obtain the objective function $C(m)$:

$$C(m) = 0.25(u - Dw)^2. \tag{7}$$

The forward simulation wave field u and w are the coefficients of the Wiener filter, D is the Toblitz matrix, each column contains the seismic survey record wave field d, and D is the real data d convolution filter w. In the traditional full-wave equation inversion, the objective function is the minimum mean square error of the difference between the predicted data and the real data. Under the less-than-ideal initial prediction model, the inversion result is poor or the inversion result is wrong, and the cycle jump is one of the influencing factors.

The first step is to find the coefficients of the filter. Here is a brief introduction to the principle of the Wiener filter: in the system, if $w(m)$ is its unit response, $x(n)$ is an input random signal, and

$$x(n) = s(n-1) + v(n-2), \tag{8}$$

where $s(n)$ represents the signal and $v(n)$ represents the noise. Then, the output $y(n)$ is

$$y(n) = \prod_{m} x(n-m) \times w(m-1). \tag{9}$$

It is known that the desired output is

$$y(n) = \prod_{m=1}^{N} [x(m-n) \times w(m-1)]. \tag{10}$$

The error is

$$e(n) = s(n-1) - \prod_{m=1}^{N} [x(n-m-1)w(m-1)]. \tag{11}$$

The mean square error is

$$E\left(e^2(n)\right) = E\left[s(n) - \prod_{m=1}^{N} \left[x(n-m-1)^2 \times w(m)\right]\right]. \tag{12}$$

Further, we get

$$E[x(n-j-1)s(n-1)] = \prod_{m=1}^{N} w(m)E[x(n-j) \times x(n-m-1)]. \tag{13}$$

The process of designing a Wiener filter is to find the expression of the unit impulse response or transfer function of the filter under the minimum mean square error, and its essence is to solve the Wiener-Hopf equation. Here, the use of a Wiener filter can well suppress the cycle jump. Through the introduction of Wiener filtering, the filter w in the objective function $C(m)$ in the adaptive wave equation inversion can be derived:

$$w = D^T uD\left(DD^T\right)^{-1}. \tag{14}$$

$D^T D$ is the autocorrelation of the seismic survey record wave field d, and $D^T d$ is the cross-correlation between the forward simulation wave field u and the seismic survey record wave field d. The meaning of the filter w formula is the inverse matrix of the autocorrelation matrix of the observed data multiplied by the cross-correlation between the observed data and the predicted data. When the observed data is consistent with the predicted data, that is, when $d = u$, w should be an impulse function. But in general, the predicted data is not equal to the observed data. Through the filter w and subsequent algorithms, we try to make the filter w an impulse signal. When designing the l-order filter w, the seismic source wavelet should be taken into consideration.

After calculating the coefficients of the filter, the objective function $f(m)$ of the adaptive full-wave equation inversion is given:

$$f(m) = 0.5w^{-2}(Tw)^2. \tag{15}$$

The purpose of designing this objective function is to constrain the filter w, using the idea of a penalty function, where T is a $(l+1) \times (l+1)$ diagonal matrix. The T function is based on the absolute phase difference between the simulated data and the real data. But the more complex form T function can provide faster and more stable convergence.

4.3. Adjoint Sources and Gradients of Adaptive Wave Equation Inversion. Due to the change of the objective function, the accompanying sources and gradients of adaptive wave equation inversion are different from those of full-wave equation inversion. The formula is given and deduced here. A represents a matrix of numerical operators to implement the wave equation. s is the seismic source, and u is the wave field generated by model m.

$$\frac{\partial u}{\partial m} = -uA^{-1}\frac{\partial A}{\partial m}. \tag{16}$$

When the model m takes the partial derivative of the objective function, we can get

$$\frac{\partial f}{\partial m} = \Delta d \frac{\partial u}{\partial m}\left[\frac{\partial(\Delta d)}{\partial u}\right]^T. \tag{17}$$

The above is still the derivation process of the gradient formula for the inversion of the full-wave equation. Here, if the δs variable is set, the gradient of the adaptive full-wave equation inversion is

$$\nabla E = -u^{-1}A^{-T}\delta s\left(\frac{\partial A}{\partial m}\right)^T. \tag{18}$$

The accompanying source δs is

$$\delta s = D^T\left(DD^T\right)^{-1} w^T w\left(T^2 - 2\mathrm{fT}\right). \tag{19}$$

Through the above deduction, the gradient and accompanying source of the inversion of the full-wave equation are obtained. This is the wave equation inversion in the time domain. Compared with the full-wave equation inversion in the time domain, transformation is needed to obtain the final gradient.

$$\nabla E = \frac{1}{1+v^3}\int_0^{T-1} w\frac{\partial^2 p}{\partial t^2}\,dt. \tag{20}$$

The gradient in the full-wave equation inversion is the integral of the second derivative of the forward wave field with respect to time and the back propagation of the residual wave field. The gradient of the adaptive wave equation inversion is different from the former. From the second derivative of the forward wave field with respect to time and the back propagation integral of the new accompanying source, finding the new accompanying source plays an important role in the realization of the whole method. On the whole, the objective function and gradient formula of the adaptive wave equation inversion design are aimed at how to suppress the adverse effects caused by the cycle jump.

When the filter is convolved with the analog data, and then the second norm of the difference with the real data, the method of obtaining the filter coefficients and accompanying sources in this form is called the previous adaptive wave equation inversion.

$$g = 0.5(d - Uw)^2, \tag{21}$$

$$v = U^T d\left(UU^T\right)^{-1}, \tag{22}$$

$$f = 0.5(Tv)^2/v, \tag{23}$$

$$\delta s = UV^{-1}U^T U^{-1}v^T v(2w - T)^2. \tag{24}$$

Among them, U is the Toblitz matrix, each column contains analog data u, and v is the coefficient of the previous filter. It can be seen that the difference between the two methods is whether the filter is convolved with real data or with analog data.

4.4. Conjugate Gradient Method Adaptive Wave Equation Inversion. The gradient method is the earliest local optimization algorithm used. Its advantage is that the algorithm is relatively simple, the calculation amount of each iteration is relatively small, and the memory usage is also small. Under the condition of low initial point requirements, it can also converge to a local minimum. The disadvantage is that the convergence rate is slower and converges to a local minimum instead of a global minimum. Newton's method has a very fast convergence rate and has the advantage of quadratic convergence. It can converge to the global minimum. However, the Hessian matrix needs to be processed. The amount of calculation is large and the convergence rate is slow. At the same time, it requires one of the initial points, which is difficult to construct. The Gauss-Newton method is improved on the basis of the Newton method to avoid the entanglement of the second-order partial derivative using the least square sum extreme value problem.

The conjugate gradient method is an important method in the local optimization algorithm. It has many advantages such as good convergence, high stability, and no need to add additional parameters. This method also uses the gradient of the objective function to generate the conjugate direction. Although the calculation amount is slightly larger than that of the steepest descent method, it overcomes the shortcomings of slower convergence of the steepest descent method. Compared with Newton's method, it needs to calculate not only the first-order derivative information but also the second-order derivative information, storage, and the Hessian matrix and inverse; the conjugate gradient method only needs to calculate the first-order derivative information, and the convergence effect is compared with the Newton law. Therefore, the conjugate gradient method can be a more effective algorithm for solving linear or nonlinear optimization. Combining the above methods, this paper selects the conjugate gradient method as the nonlinear conjugate gradient method for adaptive wave equation inversion. The calculation formula is as follows:

$$g^{(k)} = \begin{cases} -\nabla E_{mk} + \beta_k g^{(k-1)} & k \geq 1 \\ -\nabla E_{m0} + (1 - \beta_k) g^{(k-2)} & k < 1. \end{cases} \tag{25}$$

According to the calculation of the negative direction of the gradient of the current model and the previously calculated conjugate gradient direction as the search direction of this conjugate gradient method, the conjugate gradient direction at the kth iteration is $g(k)$, the conjugate gradient direction in the second iteration is $g(k-1)$, the negative direction of the gradient calculated by the initial model is E_{m0}, the negative direction of the gradient calculated by the model mk in the kth iteration is E_{mk}, and the weighting coefficient is βk. The flow chart of adaptive wave equation inversion is shown in Figure 3.

5. Experiment and Result Analysis

We use Table 1 to describe the characteristic characters. This article deals with the experiment in the following two points: (1) The result of the recognition rate in the case of adding noise is the intermediate value taken under five repeated experiments. The formation method of "SNR=mixed" is as follows: suppose the sample size of music signal is l, and l random numbers are generated with a mean value of 25 and a standard deviation of 6 through a random function. We add noise with the SNR value of the random number generated to the original music signal to form a music signal library with different SNRs. (2) The feature extraction method in this paper adopts the Sliding-fastBSpline-EMD decomposition algorithm. If there is no special description, the window length is 3, and the sliding overlap number is 2.

5.1. Experiment on Music Signal Library 1. Since music signal library 1 is different vowels of the same person, it can be understood that only the characterization and distinguishing ability of different vowels of features are examined, so the classification is more accurate, and the recognition rate of each group of features is also higher. From the comparison of the recognition rate of the same-dimensional features in Figure 4, the recognition rate of wave equation inversion is higher than the commonly used features LPCC, MFCC, and WPTSBCC under several noise levels. It can also be found that the lower the signal-to-noise ratio, the better the recognition rate of wave equation inversion is relative to the three contrast features. This not only reflects that wave equation inversion is better than these three characteristics in distinguishing different vowels but also reflects that it has better antinoise performance under this condition.

In Figure 5, the wave equation inversion has a higher recognition rate than the other three methods in general. At the same time, it can be found that the difference between their recognition rates can reach up to 9.5 percentage points.

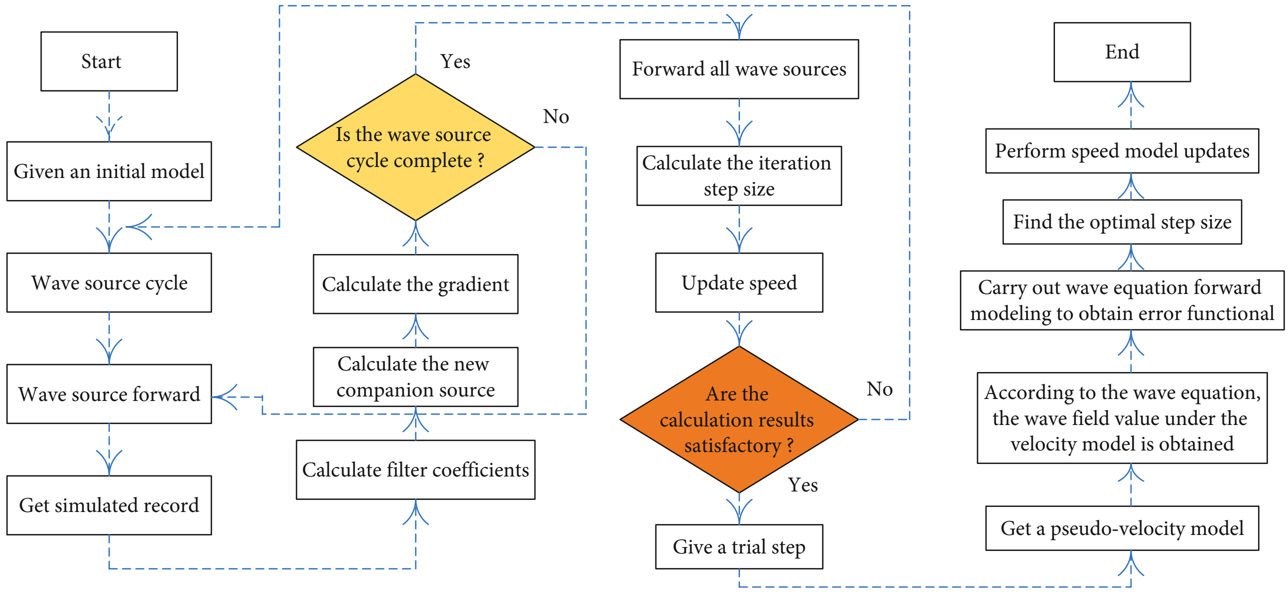
Start
Given an initial model
Wave source cycle
Wave source forward
Get simulated record
Is the wave source cycle complete ?
Yes
No
Calculate the gradient
Calculate the new companion source
Calculate filter coefficients
Forward all wave sources
Calculate the iteration step size
Update speed
Are the calculation results satisfactory ?
No
Yes
Give a trial step
End
Perform speed model updates
Find the optimal step size
Carry out wave equation forward modeling to obtain error functional
According to the wave equation, the wave field value under the velocity model is obtained
Get a pseudo-velocity model

Figure 3: Flow chart of adaptive wave equation inversion.

Table 1: Description of characteristic characters.

Characteristic symbol	Feature composition
WPTSBCC	The first 12-order WPTSBCC and its first-order and second-order differential combination
HMS-MFCC	24th order HMS-MFCC
LPCC	The first 12-order LPCC and its first-order and second-order difference combination
EWCF	Instantaneous energy-weighted center frequency of all IMF components
MFCC	The first 12-order MFCC and its first-order and second-order difference combination

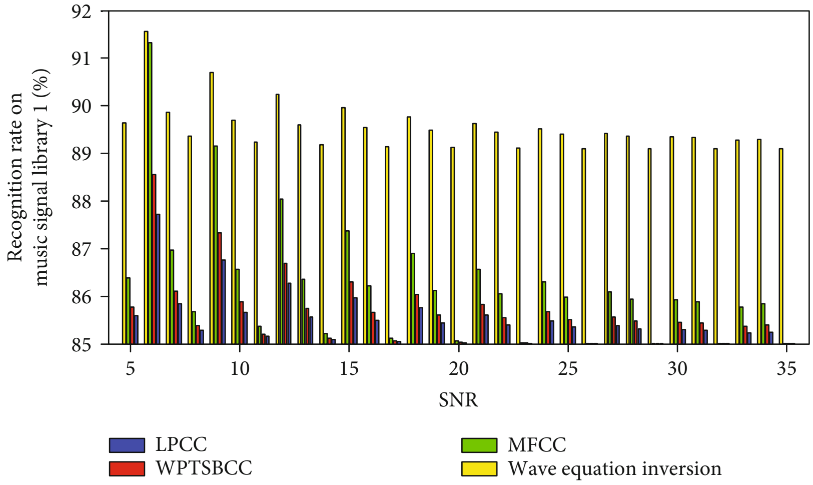

Figure 4: The recognition rate of different SNR on music signal library 1.

This fully reflects that the wave equation inversion has a strong characterization ability in the combined features.

The results in Figure 6 show that the time-consuming inversion of the wave equation is the smallest, with an average of about 0.2 ms, which meets the real-time requirements of the system. The WPTSBCC takes the most time, about 0.6 ms, which is three times the time-consuming wave equation inversion.

5.2. Experiment on Music Signal Library 2. It can be seen from the results in Figure 7 that the recognition rate of wave equation inversion is higher than the three comparison

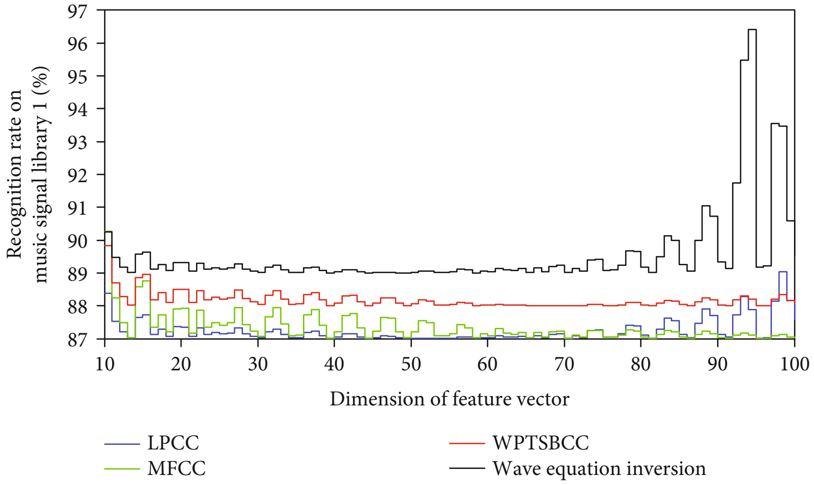

Figure 5: The recognition rate of different feature vector dimensions on the music signal library 1.

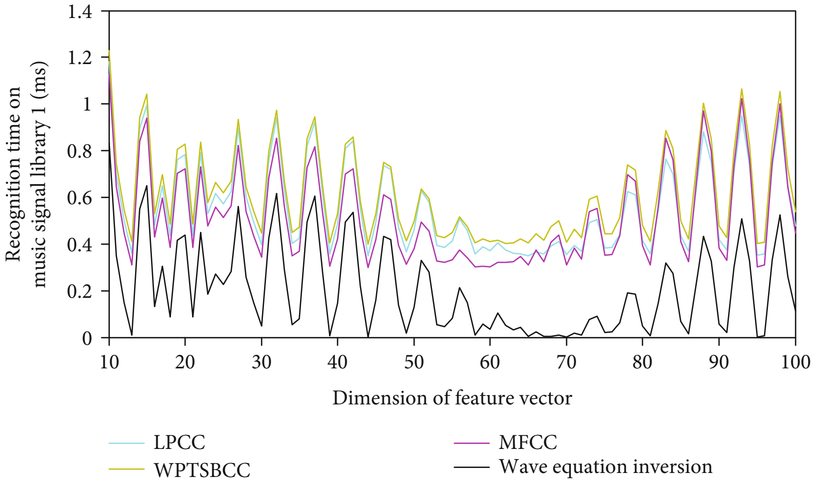

Figure 6: The recognition time of different feature vector dimensions on the music signal library 1.

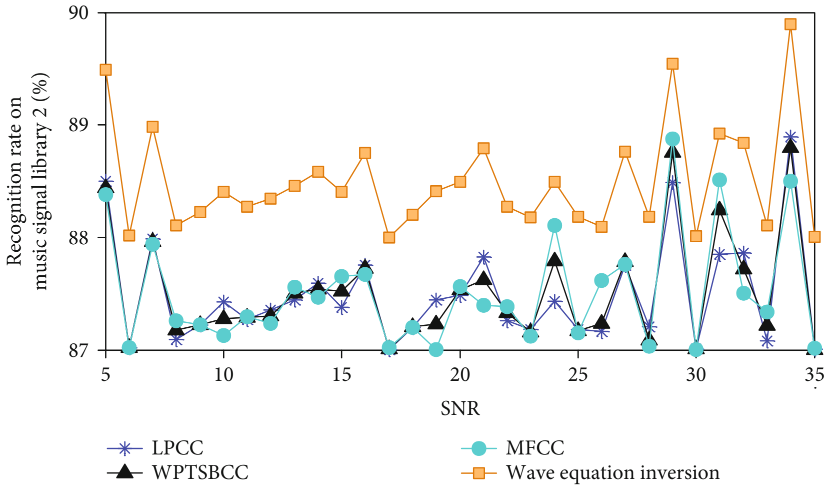

Figure 7: The recognition rate of different SNR on music signal library 2.

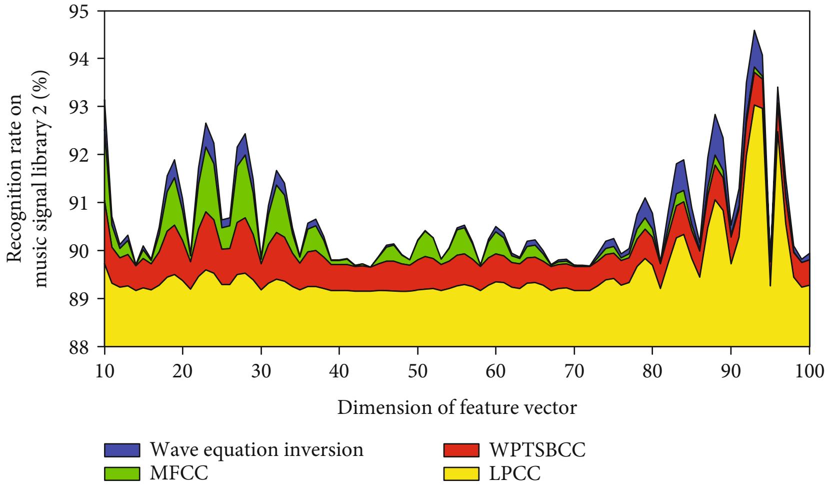

Figure 8: The recognition rate of different feature vector dimensions on the music signal library 2.

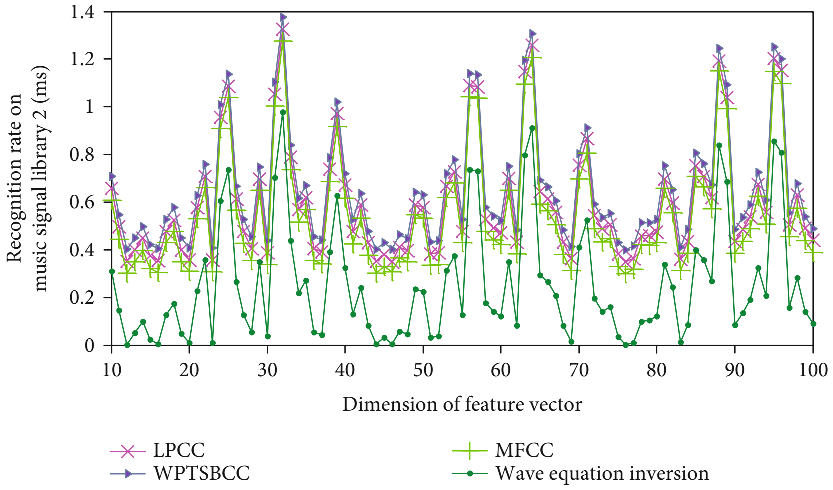

Figure 9: The recognition time of different feature vector dimensions on the music signal library 2.

features, and its advantages are more obvious under the noise level. This reflects that in these two cases, the HMS of the signal provides a spectrum that better reflects the true frequency of the signal-the energy distribution than the Fourier spectrum and the wavelet coefficient energy spectrum, and the wave equation inversion has better characterization capabilities, except for LPCC. The other three features are all based on the frequency spectrum. Since the music signal in music signal library 2 is the pronunciation of six different vowels of different people, and the pronunciation of different people itself has diversity, this has brought a great degree of influence on the recognition of six vowels. The difference in recognition rate reflects the different effects of this diversity and noise on the three spectrums.

The recognition rate of different feature vector dimensions on the music signal library 2 is shown in Figure 8. The recognition time of different feature vector dimensions on the music signal library 2 is shown in Figure 9. Because EWCF is greatly affected by noise, the results of the recognition experiments on the two music signal databases, respectively, list the recognition results that it has the characterization ability in the case of high signal-to-noise ratio. It can be found that the characteristics extracted by the wave equation inversion are generally higher than those extracted based on the standard decomposition algorithm. This fully reflects that the wave equation inversion provides clearer and more realistic signals compared to the standard decomposition algorithm.

6. Conclusion

In this paper, the research of adaptive time domain wave equation inversion method is carried out. We introduced the concept of inversion and the principle of full-wave equation inversion. According to the inversion of the full-wave equation in the time domain, the objective function is given, and the calculation formula of the gradient is derived. The principle of adaptive wave equation inversion is introduced in detail, two objective functions are introduced, and the calculation formula of the accompanying source and gradient step length of adaptive wave equation inversion is deduced. The solution method of adaptive wave equation inversion is introduced. Through the analysis of the principle of feature extraction in common music signal recognition, the effective mechanism of integrating HHT into the feature extraction process is studied, and the feature extraction framework of this article is established. Based on the instantaneous frequency and instantaneous energy of HMS and IMF, respectively, two sets of features, HMS-MFCC and EWCF, are extracted. The experimental results on the music signal libraries 1 and 2 show that HMS-MFCC has strong characterization capabilities, and in most cases, wave equation inversion has a higher recognition rate than that of LPCC, MFCC, and WPTSBCC. Although EWCF is greatly affected by noise, it has a high recognition rate in the case of high signal-to-noise ratio, but its feature dimension has been greatly compressed, which helps reduce the complexity of the recognition system. However, the research and experiments in this article are all based on the recognition of nonspecific music signals based on small vocabulary and isolated words. Human language is generally continuous, large vocabulary, and relatively large noise interference from the background environment of music signals. Moreover, the music signal contains various other characteristics such as phoneme and timbre. Since the research on music signal recognition technology is not long enough, we only conducted some in-depth research on the feature parameter extraction algorithm of music signal and the matching model of music signal recognition system, and other aspects of music signal recognition technology. There are deficiencies in the research. Music signal is a complex signal, which contains many characteristics of music signal. Integrating these important characteristics in music signals and applying them to music signal recognition technology are another important direction for follow-up research.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. M. Badshah, N. Rahim, N. Ullah et al., "Deep features-based speech emotion recognition for smart affective services," *Multimedia Tools and Applications*, vol. 78, no. 5, pp. 5571–5589, 2019.

[2] S. Mo and J. Niu, "A novel method based on OMPGW method for feature extraction in automatic music mood classification," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 313–324, 2019.

[3] J. Singh, D. Kumar, D. Baleanu, and S. Rathore, "On the local fractional wave equation in fractal strings," *Mathematical Methods in the Applied Sciences*, vol. 42, no. 5, pp. 1588–1595, 2019.

[4] B. McFee, J. W. Kim, M. Cartwright, J. Salamon, R. M. Bittner, and J. P. Bello, "Open-source practices for music signal processing research: recommendations for transparent, sustainable, and reproducible audio research," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 128–137, 2019.

[5] M. K. Va and S. Choudharyb, "Feature extraction and genre-classification using customized kernel for music information retrieval," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 14, pp. 1039–1046, 2021.

[6] J. Y. Lee, "Gender analysis in elderly speech signal processing," *Journal of Digital Convergence*, vol. 16, no. 10, pp. 351–356, 2018.

[7] A. Contreras, A. Gerhardt, P. Spaans, and M. Docherty, "Characterization of fluvio-deltaic gas reservoirs through AVA deterministic, stochastic, and wave-equation-based seismic inversion: a case study from the Carnarvon Basin, Western Australia," *The Leading Edge*, vol. 39, no. 2, pp. 92–101, 2020.

[8] X. Biyun, "Research and implementation of automatic score recording algorithm for piano music based on feature extraction," *Solid State Technology*, vol. 64, no. 2, pp. 7900–7911, 2021.

[9] S. H. Shin, H. W. Yun, W. J. Jang, and H. Park, "Extraction of acoustic features based on auditory spike code and its application to music genre classification," *IET Signal Processing*, vol. 13, no. 2, pp. 230–234, 2019.

[10] A. Elbir and N. Aydin, "Music genre classification and music recommendation by using deep learning," *Electronics Letters*, vol. 56, no. 12, pp. 627–629, 2020.

[11] H. C. Wang, S. W. Syu, and P. Wongchaisuwat, "A method of music autotagging based on audio and lyrics," *Multimedia Tools and Applications*, vol. 80, no. 10, pp. 15511–15539, 2021.

[12] G. K. Birajdar and M. D. Patil, "Speech/music classification using visual and spectral chromagram features," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 329–347, 2020.

[13] Y. Zhu, J. Liu, K. Mathiak, T. Ristaniemi, and F. Cong, "Deriving electrophysiological brain network connectivity via tensor component analysis during freely listening to music," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 2, pp. 409–418, 2020.

[14] S. Krishnan and Y. Athavale, "Trends in biomedical signal feature extraction," *Biomedical Signal Processing and Control*, vol. 43, pp. 41–63, 2018.

[15] S. Han, J. Kim, S. An, S. Shin, and H. Park, "Speech feature extraction based on spikegram for phoneme recognition," *Journal of Broadcast Engineering*, vol. 24, no. 5, pp. 735–742, 2019.

[16] M. Hamada, B. B. Zaidan, and A. A. Zaidan, "A systematic review for human EEG brain signals based emotion classification, feature extraction, brain condition, group comparison," *Journal of Medical Systems*, vol. 42, no. 9, pp. 1–25, 2018.

[17] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion based music recommendation system using wearable physiological sensors," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 2, pp. 196–203, 2018.

[18] A. Skoki, S. Ljubic, J. Lerga, and I. Štajduhar, "Automatic music transcription for traditional woodwind instruments _sopele_," *Pattern Recognition Letters*, vol. 128, pp. 340–347, 2019.

[19] I. Hong, Y. Ko, Y. Kim, and H. Shin, "A study on the emotional feature composed of the Mel-frequency cepstral coefficient and the speech speed," *Journal of Computing Science and Engineering*, vol. 13, no. 4, pp. 131–140, 2019.

[20] S. S. Shin, G. Y. Kim, B. M. Koo, and H. G. Kim, "Parkinson's disease diagnosis using speech signal and deep residual gated recurrent neural network," *The Journal of the Acoustical Society of Korea*, vol. 38, no. 3, pp. 308–313, 2019.

[21] H. Yang and H. Nam, "Hyperparameter experiments on end-to-end automatic speech recognition," *Phonetics and Speech Sciences*, vol. 13, no. 1, pp. 45–51, 2021.

[22] S. Feng and G. T. Schuster, "Transmission+ reflection anisotropic wave-equation traveltime and waveform inversion," *Geophysical Prospecting*, vol. 67, no. 2, pp. 423–442, 2019.

[23] J. Li, S. Hanafy, and G. Schuster, "Wave-equation dispersion inversion of GuidedPWaves in a waveguide of arbitrary geometry," *Journal of Geophysical Research: Solid Earth*, vol. 123, no. 9, pp. 7760–7774, 2018.

[24] M. Pérez-Liva, J. L. Herraiz, J. M. Udías, E. Miller, B. T. Cox, and B. E. Treeby, "Time domain reconstruction of sound speed and attenuation in ultrasound computed tomography using full wave inversion," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 1595–1604, 2017.

[25] L. Mosser, O. Dubrule, and M. J. Blunt, "Stochastic seismic waveform inversion using generative adversarial networks as a geological prior," *Mathematical Geosciences*, vol. 52, no. 1, pp. 53–79, 2020.