

## Research Article

# Music Signal Recognition Based on the Mathematical and Physical Equation Inversion Method

Wei Jiang<sup>1</sup> and Dong Sun<sup>2</sup>

<sup>1</sup>Department of Music, Shandong University of Science and Technology, Qingdao 266590, China

<sup>2</sup>School of Electronic and Information Engineering, Henan Institute of Technology, Xinxiang 453003, China

Correspondence should be addressed to Wei Jiang; [skd992977@sdust.edu.cn](mailto:skd992977@sdust.edu.cn)

Received 24 August 2021; Accepted 20 September 2021; Published 1 October 2021

Academic Editor: Miaocho Chen

Copyright © 2021 Wei Jiang and Dong Sun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Digitization and analysis processing technology of music signals is the core of digital music technology. The paper studies the music signal feature recognition technology based on the mathematical equation inversion method, which is aimed at designing a method that can help music learners in music learning and music composition. The paper firstly studies the modeling of music signal and its analysis and processing algorithm, combining the four elements of music sound, analyzing and extracting the characteristic parameters of notes, and establishing the mathematical model of single note signal and music score signal. The single note recognition algorithm is studied to extract the Mel frequency cepstrum coefficient of the signal and improve the DTW algorithm to achieve single note recognition. Based on the implementation of the single note algorithm, we combine the note temporal segmentation method based on the energy-entropy ratio to segment the music score into single note sequences to realize the music score recognition. The paper then goes on to study the music synthesis algorithm and perform simulations. The benchmark model demonstrates the positive correlation of pitch features on recognition through comparative experiments and explores the number of harmonics that should be attended to when recognizing different instruments. The attention network-based classification model draws on the properties of human auditory attention to improve the recognition scores of the main playing instruments and the overall recognition accuracy of all instruments. The two-stage classification model is divided into a first-stage classification model and a second-stage classification model, and the second-stage classification model consists of three residual networks, which are trained separately to specifically identify strings, winds, and percussions. This method has the highest recognition score and overall accuracy.

## 1. Introduction

The emergence of computer technology and the Internet has facilitated the birth and development of a series of interdisciplinary disciplines that combine science and art. In the field of music research, music, as an artistic discipline closely connected with daily life and learning, is gradually going digital and technological. In recent years, modern music technology, especially electronic music technology, has made rapid development, and issues such as music recognition, retrieval, and synthesis based on computer technology have received more attention from researchers [1]. Traditional music teaching requires professional teachers to tutor students, and teaching is characterized by repetitive practice. This repetitive work

not only greatly reduces the effective utilization of teachers, but also charges expensive fees for one-on-one teaching and tutoring, making systematic music learning impossible for families with low-income levels. In addition, in the teaching process, musicians judge pitch based on their rich teaching experience, based on what the human ear hears, which is too subjective and less accurate to make mistakes [2]. If computer technology is applied to music teaching, on the one hand, it can assist musicians in music teaching to reduce labor intensity, and on the other hand, music learners can carry out music learning independently from teachers to a certain extent and reduce learning costs. In addition to playing a significant role in music teaching, digital music technology can also promote the development of intelligent music

composition [3]. The realization of music synthesis technology makes automatic music composition possible, and for people who are not very proficient in music theory, music synthesis technology lowers the threshold of music composition, so that more music lovers can create their works. In addition, music synthesis technology also contributes to the development of electronic instruments and the improvement of the sound of traditional instruments [4].

With the development of artificial intelligence technologies, music information retrieval has received renewed attention in the field of computer science. Content-based music information retrieval includes several research directions: music recognition, melody extraction, pitch estimation, sentiment classification, rhythm detection, genre, and style classification [5]. Among them, the identification of multiple instruments in a music song and the prediction of their activity levels is an important research topic in the MIR task. Music recognition techniques can be applied in many contexts, such as searching for songs with specific instruments or identifying the starting and ending positions of a certain instrument played in the audio. Modeling the performance of music recommendation systems based on user preferences for certain instruments can be improved. They can also be used for automatic music transcription in polyphonic music, playback technique detection, and source separation tasks, where preconditioning models for the specific instruments present in the source separation task may improve its performance [6]. Multimusic recognition is essentially a timbre perception task. The tone is a subjective property that is difficult to quantify. A person with good musical sense and professional training can easily identify instruments in audio. However, the vast amount of music cannot rely on a human to identify and then provide labeled information for retrieval. With the development of artificial intelligence and computing power, we can extract the corresponding features of musical instruments in audio files and train efficient deep convolutional networks to achieve automatic recognition of musical instruments.

Music signals, as a type of audio signal, are widely distributed through the convenient Internet. With the permission of the copyright, people can download various kinds of music on the Internet. Therefore, the data volume of music audio is getting larger and larger, and the requirements for the retrieval task are getting higher and higher. However, many mainstream music search engines are still based on simple text retrieval, manually labeled song titles, artists, or years. It would be significant for retrieval efficiency and user experience if retrieval could be based on the content information of the music signal itself, and these features could be automatically identified. Chapter 1: Introduction. Firstly, the background and significance of the paper are explained in the context of the current social situation and social needs, and the main research content and the arrangement of each chapter are given. Chapter 2: Related Work. A research analysis of the current research method is conducted, and some knowledge of music theory and the basic elements of digital audio are introduced, which are conducive to an in-depth understanding of the essential characteristics of musical instruments and the key features of

construction identification. Chapter 3: Research on Music Signal Recognition Based on Mathematical Equation Inversion Methods. In terms of recognition, the paper chooses to characterize the original signal using Mel inversion coefficients. Then, the single note recognition algorithm is introduced, and based on it, the note cutting algorithm is studied to achieve multinote recognition. In terms of synthesis, mathematical modeling of the music signal is studied, and additive synthesis techniques are used to achieve piano tone reproduction based on the music score as well as note time value information. Chapter 4: Analysis of Results. Chapter 5: Conclusion. It mainly summarizes the final research results of the paper, analyzes the shortcomings of the paper in the research process, and also provides an outlook for future work because of these shortcomings.

## 2. Related Work

Douglas Nunn proposed a music recognition system based on the inverse signal processing method of mathematical equations. The maximum number of articulations that the system can recognize is increased to 8, but the accuracy of the system is not very high because it is more concerned with the consistency of the recognition results with the auditory perception [7]. Since the inverse mathematical equation approach network uses a distributed collaborative approach to eliminate the global control module, researchers began to apply the inverse mathematical equation approach network to music recognition systems. [8] The successful application of Bayesian networks in music recognition systems has proven to result in better a priori knowledge of the system. In recent years, researchers have started to apply fuzzy neural networks to music recognition. It has been verified that this method is closest to the human cognitive process of music and can effectively extract music information; thus, it has been more widely used. Ambrosanio et al. proposed an automatic music emotion recognition method based on a mathematical equation inversion model and gene expression programming algorithm, which has a high recognition rate for single-emotion music, but poor recognition for complex music with multiple emotions [9]. Yatabe et al. applied the tonal level contour feature in the chord recognition algorithm and achieved satisfactory recognition results [10].

In monophonic music, it has been possible to perform musical recognition of audio fragments at the note level or continuous audio signals played by solo instruments [11]. He et al. proposed a linear spectral feature that was used together with a Gaussian mixture model to evaluate the classification of instrument families and to classify instruments by 14 instrument families. In addition to analyzing predefined features for classification, a classifier can be used to learn the features to accomplish the classification task [12]. Long et al. used sparse spectral coding and support vector machines to classify single and multisource audio [13]. Jiang et al. proposed to extract the Meier spectrogram from a dataset of single note clips of 24 musical instruments, use sparse coding to learn the features in the spectrogram, and then train a support vector machine to use the learned features to classify the instruments with an accuracy of about 0.95

for the 24 instrument categories [14]. The deep architecture allows for end-to-end training of the feature extraction and classification modules to “learn” features, resulting in higher accuracy than traditional methods [15]. The successful application of deep learning in these two scenarios, based on monophonic music recognition and major music recognition in polyphonic music, inspires us to further apply it to polyphonic music recognition [16].

To perform multimusic recognition in polyphonic music, general time-frequency features may not achieve good recognition, so we selected pitch features and mathematical equation inversion recognition as the input features of the model. The pitch features reflect the range of the instrument and the fundamental frequency of the notes, and we use the idea of multibasis frequency estimation to extract the pitch features by using a filter set with custom parameters to extract the initial features of the audio fed to the convolutional network. Numerical equation inverse identification is a special wavelet transform that has been improved to facilitate music analysis, which can reflect the energy distribution of each pitch. We use an improved fast computational method to extract the numerical equation inverse identification of the audio. These two features combined can effectively capture the harmonic structure of the music signal, which is reflected in the music as the timbre of the instrument [17, 18]. We are currently not aware of any work on correlating timbre with the pitch in music recognition. Finally, we feature-processed the extracted features and constructed three classification models, namely, a baseline model, an attention network-based classification model, and a two-level classification model. The baseline model demonstrates the effectiveness of pitch features in music recognition; the attention mechanism has been widely used in computer vision, and we apply it to the “auditory” attention of music signals. The two-level classification model first performs coarse classification of instrument families and then performs subclassification of specific instruments of the corresponding instrument families, and the hierarchical recognition is consistent with the basic. The two-level classification model first performs the coarse classification of instrument families and then the subcategorization of specific instruments of the corresponding instrument families. A series of comparative experiments of the three classification models also explore the validity of various known experiences in multimusic recognition, as well as the possibility of unknown methods.

### 3. Research on Music Signal Identification Based on Inversion Method of Mathematical Equations

*3.1. Music Signal Feature Parameter Extraction.* A complete musical work is composed of different parts; partially it is composed of different motives, sections, and phrases. As a whole, it is composed of complete sections, parts, or movements. So the musical signal has both a whole character, and it is the interaction and connection between them that constitutes the musical integrity. The overall characteristic

is expressed by the main theme of the music, and the local characteristics are developed around the overall characteristic. That is, the relationship between commonality and individuality, where the commonality determines the individuality and the individuality reflects the commonality [19]. The study of the overall characteristics and local characteristics of the musical signal can reveal the essential characteristics of the musical signal. An individual section is the smallest unit of the musical signal that can be divided, as it is already clearly expressed in the expression of musical ideas and the shaping of musical images. Therefore, in this project, we take a single section as the smallest unit of music analysis. Figure 1 shows the characteristic diagram of the music signal. The rhythm of slow music changes slowly, and the music signal is soft. Through mathematical equations, it extracts musical features such as motives, festivals, phrases, passages, music clubs, and movements.

The expression for the spectral energy is shown in Equation (1), which is a statistical quantity. The elemental representation is based on the study of the fundamental frequency cycles of human perception, a method also commonly referred to as the chromaticity vector method. In the vector, each element corresponds to one of the 12 traditional cycles [20]. The value obtained by finding the root mean square of the spectral energy is a physical quantity related to the intensity of the sound. In note modeling and single note recognition, only the note segment needs to be detected from the speech mixed with blank noise segment. Therefore, the paper selects the short-term average energy with less calculation and better real-time performance for endpoint detection.

$$E(x) = \text{avg} \left( \lim_{N \rightarrow \infty} \sum_{i=1}^N |e_i - 1|^2 \right). \quad (1)$$

Each critical spectrum expansion function has 10 dB and 25 dB expansion to both high and low frequencies, respectively. The masking effect of the low-frequency band on the high-frequency band is strong. The effect of the critical band  $x_i$  on  $x_j$  satisfies Equation (2), where  $f(x) = x_i - x_j$ . The music signal is also different from the general audio signal in that it has not only the genre division but also the song style division. From the point of view of music theory, the beat usually occurs at the point of note onset, and the selection of the frame has a direct impact on the characteristics of the signal, as the instrument is articulated and played, and the singer sings and ends according to the beat in an orderly manner [21, 22]. The speed of the beat usually represents the style of the music signal; generally speaking, the signal spectrum changes more intense music beat faster, and the music signal is more active. Softer music has slower beat changes and softer music signals.

$$Z_{\text{db}}(f(x)) = \alpha + \beta(\nabla f(x) + \chi). \quad (2)$$

The all-pole model obtained by linear predictive analysis

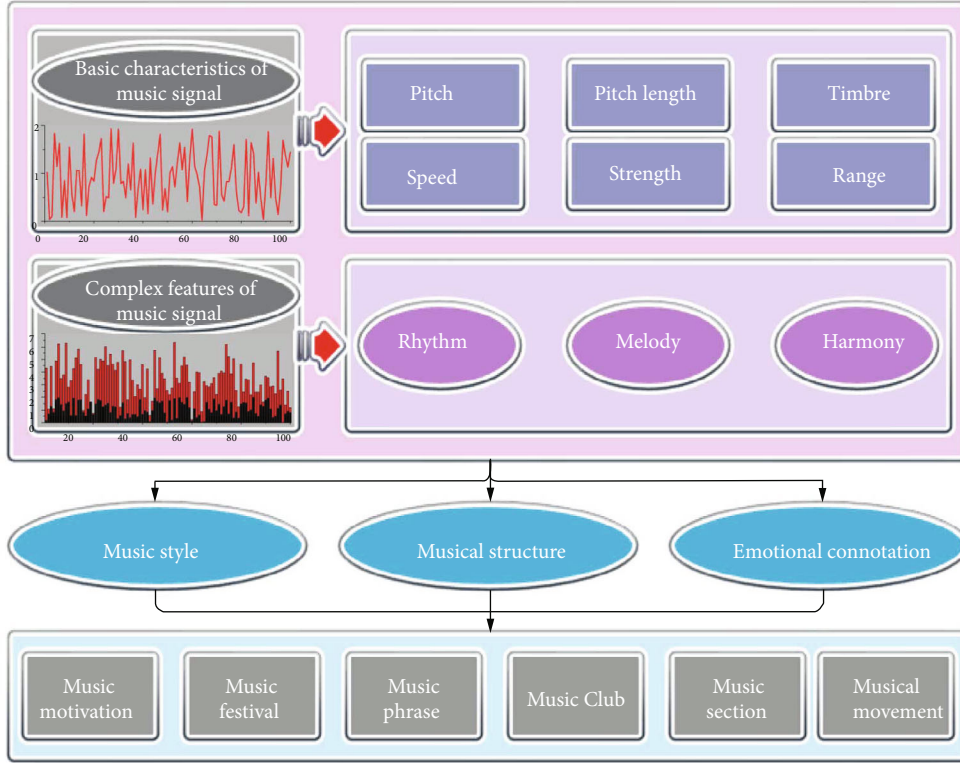


FIGURE 1: Characteristic diagram of music signal.

has a system function of Equation (3).

$$F(x) = 1 - \lim_{M \rightarrow \infty} \sum_{i=1}^M (\beta + 1) f(x_i)^{-2}. \quad (3)$$

In Equation (3),  $M$  is the order of the linear predictor. If the impulse response is assumed to be  $f(x)$ , we have Equation (4). However, since the LPC cepstral coefficients are only based on the prediction of linear relationships, the robustness of the parameters is not very good and the noise immunity is low.

$$F(x) = \lim_{M \rightarrow \infty} \sum_{i=1}^M f(x) x^{-i}. \quad (4)$$

When a speech signal is transmitted in a traveling wave across the cochlear basilar membrane, the transmission distance of the low-frequency signal is greater than that of the high-frequency signal due to its low frequency and long wavelength; thus, the high-frequency signal is masked by the low-frequency signal, and the masking ability of the higher frequency sound varies for different frequencies, and the higher the audio, the greater the masking ability [23]. Therefore, the human ear hearing system is equivalent to a filtering system to filter the treble. In terms of design implementation, a set of band-pass filters can be designed, which are arranged from dense to sparse according to the masking ability of each frequency point based on the hearing characteristics of the human ear. The conversion relation-

ship between linear frequency and  $H$  frequency is shown in Equation (5).

$$H(x) \triangleq 4096 * \ln \left( 1 + \frac{x}{900} \right). \quad (5)$$

The logarithmic energy output of each triangular filter bank is calculated as shown in Equation (6).

$$E(n) = \lg \left( \lim_{N \rightarrow \infty} \sum_{i=1}^N (e_i - 1) H(n) f(x_i) \right), n \subseteq [0, N]. \quad (6)$$

The MFCC is obtained by doing a discrete sine transformation on  $E(n)$ , and the transformation equation is as in Equation (7).  $K$  is the dimension of the characteristic parameter. Since the Mel frequency cepstrum coefficient not only responds to the human ear hearing effect but also does not make any assumptions and restrictions on the input signal, it has better robustness.

$$M(n) = \lim_{N \rightarrow \infty} \sum_{i=1}^N E(n) \sin \left( \frac{n(i - 0.25)\pi}{K} \right), n \subseteq [0, K]. \quad (7)$$

**3.2. Mathematical Equation Inversion Identification Algorithm.** In this study, we propose a procedure to calculate the adaptive crossover rate and variation rate using the population concentration by adding an extra procedure to calculate the population concentration in between the selection operation and the crossover operation. The population

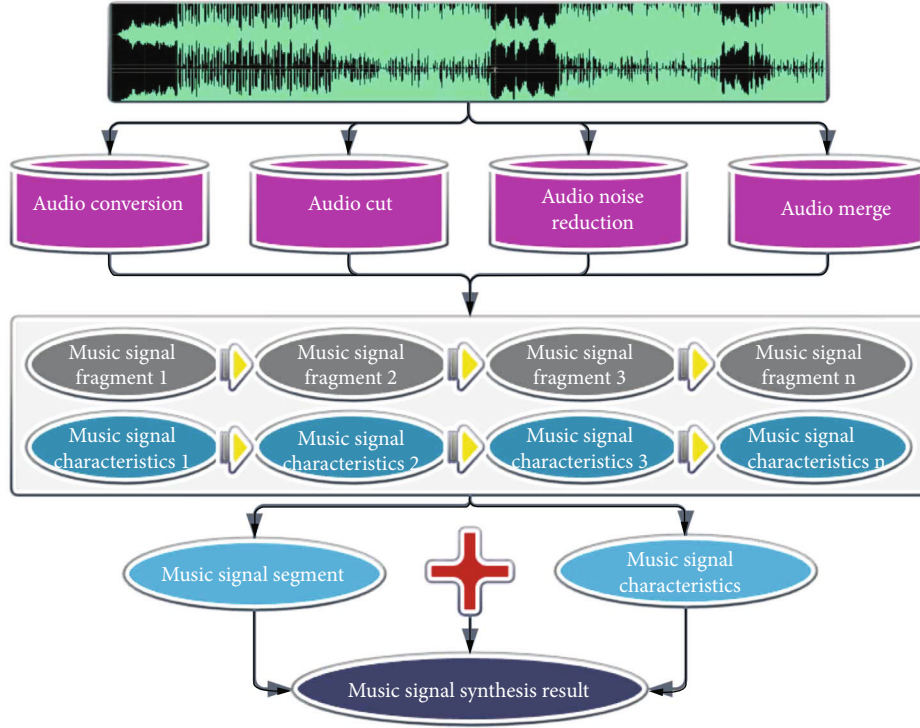


FIGURE 2: Additive synthesis schematic.

concentration  $J$ , used in this study, is calculated as Equation (8).  $N$  is the number of evolutionary generations.

$$J(N) = \frac{f(X_1(N))}{f(X_n(N))}. \quad (8)$$

Because of the great randomness when the articulator vibrates, the length of articulation time cannot be well controlled. If the linear uniform expansion method is used to align the frame lengths of the text file and the template file, it will ignore the time length transformation of each small segment in the audio file under different circumstances, leading to the result of a low recognition rate [24]. The population concentration  $J$  will be used to regulate the crossover rate and variation rate, and when the population is dispersed, the strategy of more crossover and less variation is adopted to increase the exploitation. When the population is concentrated, a strategy of more variation and less crossover is used to increase exploration. The specific settings of the adaptive crossover rate and variable rate are shown in Equation (9). The  $m_n$  can be adjusted as needed to control the crossover rate and variable rate to fluctuate in the specified range.

$$\begin{cases} f(j) = m_1 + m_2 * (J - 1), \\ f(k) = m_3 + m_4 * J. \end{cases} \quad (9)$$

The inverse algorithm is initialized by randomly generating a model  $K$  with the structure of Equation (10), saving it in the population, and setting the evolutionary algebra  $J$  to 0.

The initialization is run only once at the start of the genetic algorithm.

$$K = \{\beta_i; p_i\}, i \subseteq [1, N]. \quad (10)$$

In Equation (10),  $N$  denotes the number of layers fitted by inversion.  $\beta_i$  is the dielectric constant of layer, and  $p_i$  is the thickness of layer. The population is expressed as Equation (11).  $J$  denotes the evolutionary algebra.

$$P(J) = \|K_i(J)\|, i \subseteq [1, N]. \quad (11)$$

The music signal recognition record  $fs(t)$  of each model is compared with the measured data  $fr(t)$ , and the adaptation value  $\theta(t)$  of each model is calculated. The calculation is determined by the objective function. The objective function for this study is set as Equation (12).

$$\theta(t) = \max \left( \sqrt{|fr(t) - fs(t)|} \right). \quad (12)$$

$fr(t)$  is the measured waveform data, and  $fs(t)$  is the inverse fitted waveform data, and this objective function is to minimize the error of the measured and synthesized waveform data. At the same time, the error is also set to the adaptation value of the model, and the smaller the adaptation value of the model, the better it is. The forward and backward processes continuously interact so that the models close to the subsurface medium are retained and similar children are reproduced to eliminate the poorly fitted models. After several generations of evolution, the population model will

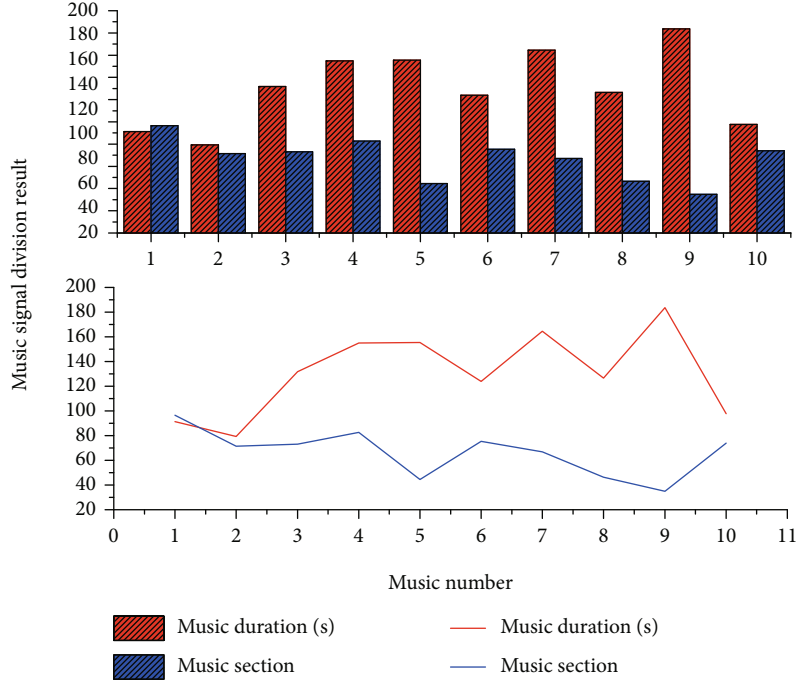


FIGURE 3: Results of different music signal divisions.

gradually approximate the measured stratigraphic model, and the optimal model of the population will be output after the evolution is completed to obtain the inversion results of the music signal.

$$f(x) = \lim_{N \rightarrow \infty} \sum_{i=1}^N |fr(t) - fs(t)|^{1/i}. \quad (13)$$

Since the search process of the optimal path is constrained by the slope, some frames cannot be matched in the actual process of solving the optimal solution. Therefore, the improved DTW algorithm takes the constraints into full consideration and reduces the matching computation between unnecessary information frames. The effective computation range of the dynamic regularization algorithm can be divided into three parts:  $(1, f(a))$ ,  $(f(a) + 1, f(b))$ , and  $(f(b) + 1, Y)$ .  $f(a)$  and  $f(b)$  take the values of the two closest integers.

$$\begin{cases} f(a) = \frac{2X - Y}{3}, \\ f(b) = 1 - \frac{2Y - X}{3}. \end{cases} \quad (14)$$

When template matching is performed, each frame on the  $x$ -axis with the parameters to be identified only needs to be compared with the frames in the interval  $[\min(y), \max(y)]$  on the  $y$ -axis, where  $\min(y)$  and  $\max(y)$  are cal-

culated as in Equation (15).

$$\begin{cases} \min(y) = 2f(a) + X - 2Y, \\ \max(y) = \frac{f(b) + 2X - Y}{2}. \end{cases} \quad (15)$$

Analytically, the range of  $\max(y)$  increases by two frames for each frame on the  $x$ -axis until  $\max(y) = X$ .  $\min(y)$  is the opposite of  $\min(y)$ , which decreases by two frames for each frame on the  $x$ -axis until  $\min(y) = 1$ . Therefore, in the actual encoding  $\min(y)$ ,  $\max(y)$ . The computational interval is obtained using Equation (16).

$$\begin{cases} \min(y) = \max(1, (X - 2) * (Y - x)), \\ \max(y) = \min(X, 2x - 1), \end{cases} \quad x \subseteq [1, Y]. \quad (16)$$

If the energy does not fluctuate much in the distribution of each frequency band, then the signal corresponding to this band of the spectrum contains more information, and the entropy value for this band of the signal is also larger. Therefore, the information entropy can be used to detect the instability of the signal and find the correct note segmentation point in the continuous notes. However, when using entropy value to segment directly, there is a problem that the audio energy is large but the spectral entropy value is small, and to solve this problem, the energy-entropy ratio is introduced. The energy-entropy ratio is the ratio of the short-time energy of each frame to the entropy value, and the spectrum of each frame is obtained by Fourier transform of the preprocessed discrete audio signal  $x(i)$ , which is given

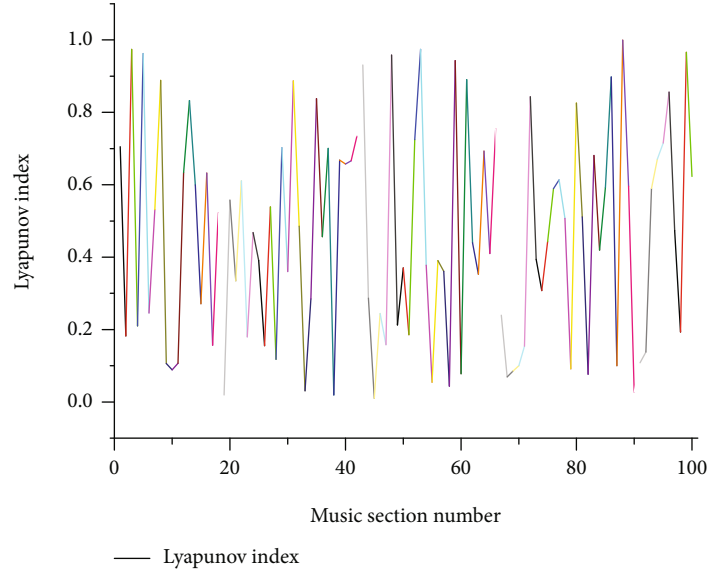


FIGURE 4: Lyapunov index line chart.

by Equation (17).

$$g(x) = \lim_{N \rightarrow \infty} \sum_{i=1}^N f(i) \ln \left( \frac{1}{x(i)} \right). \quad (17)$$

The vibration of a piano string is a set of standing wave vibrations with many overtone components, and each overtone energy is strongest during a very short period when the key is pressed and then slowly decreases to nothing over time. The High-Frequency Content (HFC) based note segmentation method uses this property of piano notes to weight the high-frequency energy in the frequency domain, thus improving the frequency domain analysis of the high-frequency band of the signal.  $H(x)$  is defined in Equation (18). The  $w(i)$  is the frequency domain weighting window, and Masri proposes to use linear weighting by  $w(i) = |i|$  for high-frequency energy.

$$H(x) = \lim_{N \rightarrow \infty} \sum_{i=1}^N w(i) |f(i)|^2. \quad (18)$$

**3.3. Music Signal Recognition Modeling.** Music synthesis is based on the analysis of musical signals, and the paper uses additive synthesis techniques in spectral synthesis to simulate musical tones generated by piano notes. The additive synthesis technique was developed from Fourier's theory that any periodic signal can be decomposed into many sinusoidal signals with different frequencies, amplitudes, and phases. Figure 2 shows the schematic diagram of the principle of additive synthesis. Define the frequency and amplitude of different harmonics and mix them together to form a new sound. But if you want to use the 1-9th harmonic to form a saw tooth-like waveform, you need an oscillator, amplifier, mixer, and thresholds to control the switch of the amplifier. Use the data equation inversion method to make the synthesizer more efficient.

The attention network-based classification model has a shortcoming that although the overall accuracy and the recognition scores of instruments with the higher frequency of occurrence are improved, the recognition of harmonic instruments is not satisfactory when the main playing instruments appear simultaneously with harmonic instruments of other instrument families. This is essentially a category imbalance problem; differences in the proportions of different categories can interfere with the learning of model parameters. When the probability of a category occurring is only 0.01, even if the model misidentifies all such categories, the error rate only increases by 0.01. This makes the model tend to get parameters that favor the recognition of a larger proportion of categories during training, while it tends to ignore a smaller proportion of categories. Some classification scenarios address this problem fundamentally by increasing the number of samples in the smaller categories, but for the multi-instrument recognition problem in music signals, category imbalance is unavoidable. This is because in the creation of various musical genres, certain instruments are suitable for melodic instruments and certain instruments are suitable for harmonic use due to their timbral characteristics and range width, and melodic instruments always appear much more frequently than harmonic instruments. We often hear various piano pieces, but rarely do we hear "trumpet pieces" or "snare drum pieces."

The two-stage classification model consists of a first-stage classification model and a second-stage classification model, which are two convolutional network models. The first-level classification model uses the inverse identification of the mathematical equation as the input feature and first coarsely classifies the instrument families in the audio signal, that is, only three coarse classification labels are available for strings, winds, and percussion. The three instrument families of strings, winds, and percussion have distinct energy characteristics. For strings, the peaks at the lower-order harmonic frequency points are distinct and sharp, and the high-

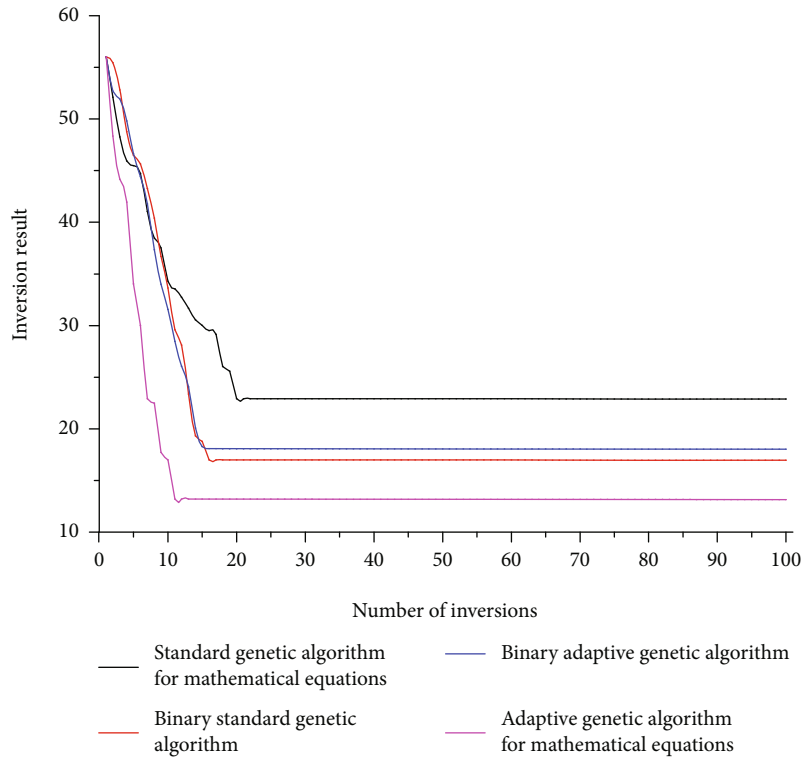


FIGURE 5: Average adaptation value curve of optimal solution per generation.

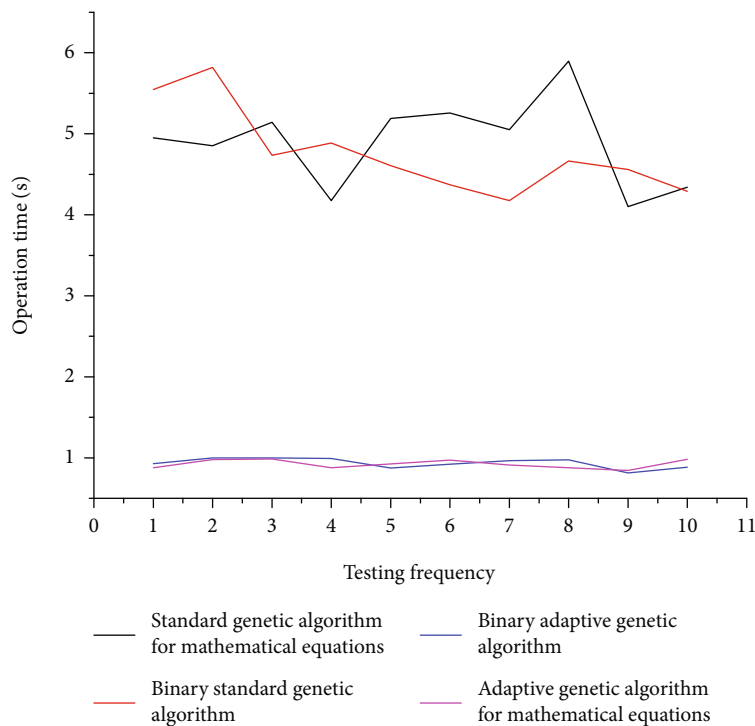


FIGURE 6: Algorithmic computing costs.

frequency harmonic amplitudes are attenuated. For wind instruments, the peaks of lower-order harmonic frequency points are sharper than those of strings, and there are still abundant harmonic spectral peaks with higher amplitude

in the high-frequency region. For percussion, the spectral peaks are not obvious, and there are also noninteger harmonics, and synthesizers often have to add white noise when synthesizing certain percussion. The inverse identification of



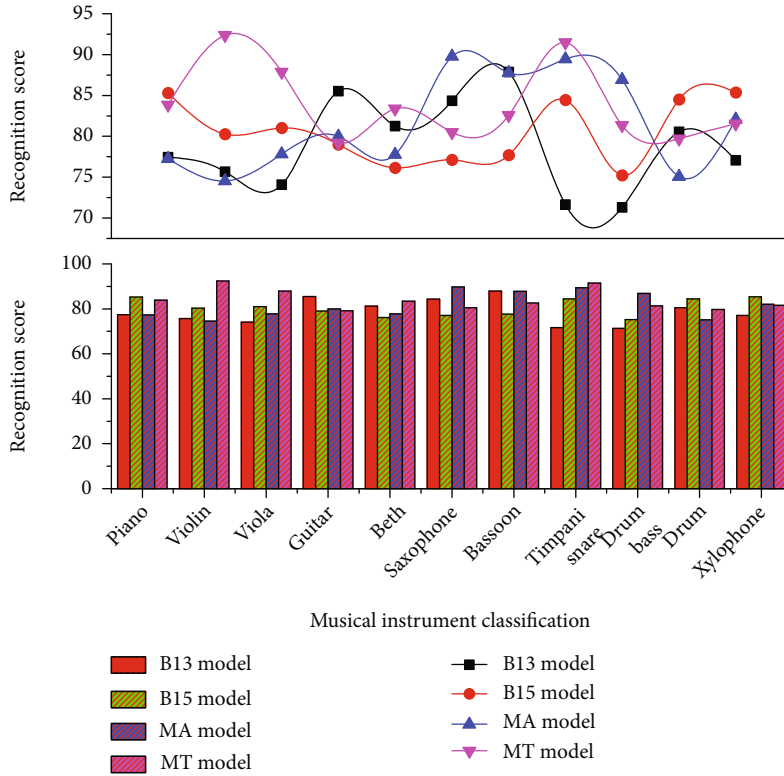


FIGURE 7: Comparison of recognition scores.

the mathematical equation reflects the time-frequency energy distribution of the audio signal, which we believe can be used as an effective feature for coarse classification. The second-level classification model consists of three residual network models with the same architecture, and each residual network model is specifically trained to identify various instruments under a certain instrument family; there is a specific network model for each of the three instrument families; based on the coarse classification results of the instrument families identified by the first-level classification model, the corresponding network models in the second-level classification model are selected, and finally, the fine classification results of each network model in the second-level classification model are selected. The subclassification results of each network model in the second classification model are aggregated as the final classification results of the audio signal.

#### 4. Analysis of Results

4.1. *Music Signal Acquisition Analysis.* According to the musical pattern, we divide the musical signal according to the bars in the pattern, and the length of the bars is determined by the length of the musical signal and the number of bars. The bars have a clear termination in the spectrum of the music signal. A total of ten different types of music signals were selected for processing in this experiment, and the results of the division of the music signals are shown in Figure 3.

Figure 4 corresponds to the statistical line graph of the results for the Lyapunov exponent of the musical signal. The Lyapunov exponent of each bar is greater than 0, indicating that each bar (local) of the piece has a chaotic character. The bar with the largest Lyapunov index is bar 92, which has the strongest chaotic feature. The smallest Lyapunov exponent is still vignette 37, which has the weakest chaotic feature. The maximum and minimum Lyapunov exponents indicate that this subsection has the largest range of chaotic features and strong nonlinear features. At the same time, the Lyapunov exponent itself is not particularly large, indicating that the musical work is a weakly chaotic system with controllable nonlinear characteristics. This is the same as the nature of musical works, where the overall trend of a musical work is controllable, but the length and intensity of a particular note at a certain moment are not precisely controllable and random.

4.2. *Music Signal Recognition Analysis.* The evolutionary efficiency analysis of genetic algorithms can be discussed in terms of the evolutionary speed and the adaptation value of the final evolutionary result. Figure 5 shows the average adaptation value curve of the optimal solution for each generation for 1000 inversions. In Figure 5, it can be seen that the evolutionary speed of the genetic algorithm with the mathematical equation coding system is significantly faster than that of the genetic algorithm with the binary system at the beginning of the evolution, but the standard genetic algorithm with the mathematical equation almost stagnates after the 10th generation and the adaptation value of the

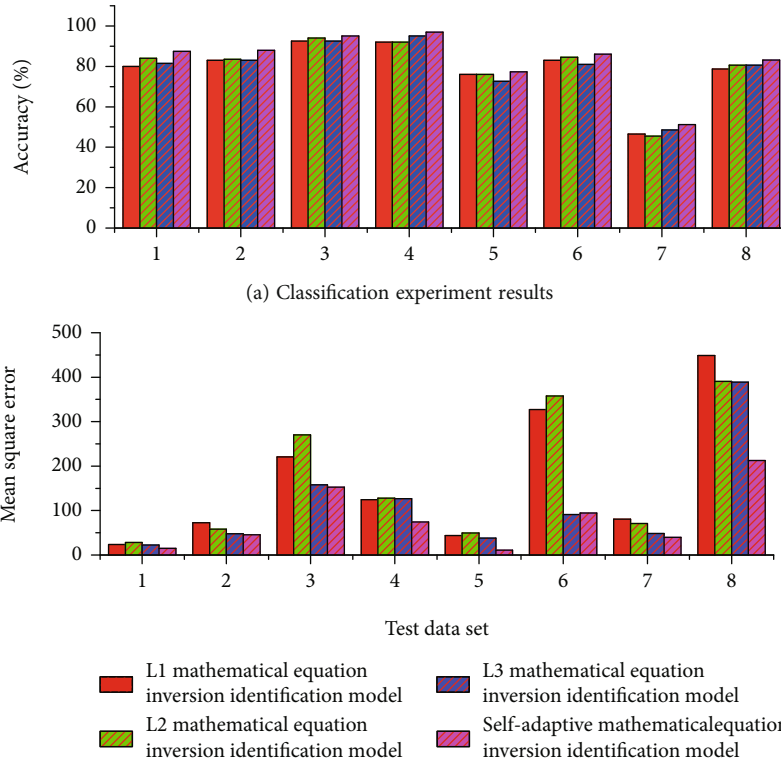


FIGURE 8: Accuracy and mean squared error results.

evolutionary result is lower than that of the binary genetic algorithm. The results of the binary standard genetic algorithm and the binary adaptive genetic algorithm are similar, but the binary adaptive genetic algorithm is faster than the mathematical equation adaptive genetic algorithm in terms of evolutionary speed. The adaptive genetic algorithm for mathematical equations is the fastest and the best in terms of evolutionary speed and evolutionary results.

The computational cost of the algorithm was analyzed by counting the computational time for each of the ten inversions, and the results are shown in Figure 6, where the platform for performing the computation is a personal computer. Since the mathematical equation coding system saves the conversion between binary and decimal, the average computing time of the genetic algorithm using the mathematical equation coding system is reduced from 4.76 s~4.89 s to 0.92 s~0.93 s, which saves 81.23%~86.16% of the computing cost. The use of the mathematical equation coding system can significantly improve the computational efficiency of wave impedance inversion.

Through the above analysis, the adaptive genetic algorithm for mathematical equations has superior performance in both evolutionary efficiency and operational efficiency. In the experiment of the music signal inversion model, the mathematical equation adaptive genetic algorithm relies on the continuous space coding system and self-adjusting crossover rate and variable rate with the evolution status, which effectively avoids the problems of poor stability and slow evolution in traditional genetic algorithm. The adaptive genetic algorithm is used for the inversion of the measured

data. It is proved that the adaptive genetic algorithm of mathematical equations has high stability and operational efficiency, and the adaptive genetic algorithm of mathematical equations is selected as the method to invert the measured data.

4.3. *Music Signal Simulation Analysis.* The experimental environment and data set threshold settings are the same as in the previous section. The inversion of the mathematical equations is used as input in the first-level classification model, and the music signal-time series matrix is output. Using the momentum algorithm with a momentum of 0.93, the minibatch size is 60, the initial learning rate is 0.05, and the weight decay factor is  $2 \times 10^{-2}$ . In the second-level classification model, we use the third-order harmonic mapping matrix I3, the fifth-order harmonic mapping matrix I5, and the sixth-order harmonic mapping matrix I6 as the input features of the string classification network, the wind classification network, and the percussion classification network, respectively. Then, the outputs of the three networks are aggregated to obtain the final music signal-time series matrix. The recognition scores and overall accuracy of various instruments in the two-level classification model can be seen in Figure 7. It can be seen that the recognition scores of most instruments are improved, especially for xylophone, which indicates that the two-level classification model alleviates the problem of category imbalance. Thus, the overall accuracy is also improved.

Using the real pitch labels and the extracted pitch features to construct harmonic mapping matrices separately

for input to the benchmark model, a comparison experiment was conducted to demonstrate that the pitch features have a positive correlation effect on multi-instrument recognition; in addition, the comparison of harmonic mapping matrices of different orders led to the conclusion that recognition of different instruments should focus on different numbers of harmonics. The attention network-based classification model, which draws on the idea of visual attention, improves the recognition scores of the main playing instruments. The two-level classification model constructs a specialized classification network for each instrument family, with the coarse classification of instrument families followed by fine classification of a specific instrument, which conforms to basic cognitive logic and alleviates the problem of category imbalance. In terms of performance, the two-level classification model has the best recognition results, and the attention network-based classification model is the most cost-effective.

In this experiment, the main models compared are the L1 mathematical equation inverse recognition model, L2 mathematical equation inverse recognition model, L3 mathematical equation inverse recognition model, and adaptive mathematical equation inverse recognition model. Figure 8(a) shows the final results of all models on classification experiments, and Figure 8(b) shows the final results of all models on regression experiments. We can see that the adaptive mathematical equation inversion recognition model achieves excellent results in terms of accuracy and mean square error, and the adaptive mathematical equation inversion recognition model can obtain higher accuracy and lower mean square error loss compared with the other three models. The general development trend of a piece of music can be inferred from the score chart, but the performance of the same piece of music cannot be the same when played by different people, and there is a lot of uncertainty in the process of performance, which does not change the overall development trend of the music. The music signal has the characteristics of chaos. By calculating the correlation dimension, we also find that the chaotic character of the music signal exists, and it remains stable at a single value despite the multiple differencing, which also shows the stability of the chaos in the music signal.

## 5. Conclusion

The system studied in the paper focuses on the application of computer science in the field of music, so to process music signals digitally, it is necessary to understand the four elements of music signals. Of these four elements, pitch and timbre are the more important characteristic parameters. From the system point of view, the music signal is a time-lagged nonlinear dynamical system, and time-lagged systems often have multiple degrees of freedom and high-dimensional characteristics. The bifurcation process is accompanied by the generation of weak chaotic phenomena. The paper firstly compares the more commonly used feature parameters in cloudy rain recognition and selects the MFCC parameter as the feature parameter for note recognition based on the comparison results. Then, the paper introduces the note recognition algorithm based on the inversion

method of mathematical equations and presents the improved DTW algorithm. In the graded recognition, we use two levels of grading, which can be increased in the future according to the number and characteristics of the recognized music. The noise in the music signal is not necessarily the AC noise that we set up. In the process of music signal data field acquisition, for example, the zero drift of the amplifier that changes with temperature, the interference around the microphone, and its instability may bring a lot of noise, so we also need to study a variety of practical situations. Considering the many different styles of music signals, we will increase the depth of research not only vertically but also horizontally in future studies to analyze many types of signals.

## 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.

## Acknowledgments

The work was supported by the Social Science Planning Project of Qingdao: The Research on the Role and Development of Music Technology in the Inheritance of Qingdao Traditional Culture and Red Culture (No.: QDSKL2001161s).

## References

- [1] M. A. Jatoti and N. Kamel, "Brain source localization using reduced EEG sensors," *Signal, Image and Video Processing*, vol. 12, no. 8, pp. 1447–1454, 2018.
- [2] G. Korvel, P. Treigys, G. Tamulevicius, J. Bernataviciene, and B. Kostek, "Analysis of 2d feature spaces for deep learning-based speech recognition," *Journal of the Audio Engineering Society*, vol. 66, no. 12, pp. 1072–1081, 2018.
- [3] Z. Wei, D. Liu, and X. Chen, "Dominant-current deep learning scheme for electrical impedance tomography," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2546–2555, 2019.
- [4] A. Massa, D. Marcantonio, X. Chen, M. Li, and M. Salucci, "DNNs as applied to electromagnetics, antennas, and propagation—a review," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 11, pp. 2225–2229, 2019.
- [5] D. C. Bowden, K. Sager, A. Fichtner, and M. Chmiel, "Connecting beamforming and kernel-based noise source inversion," *Geophysical Journal International*, vol. 224, no. 3, pp. 1607–1620, 2020.
- [6] B. Zhang, N. Bi, C. Zhang, X. Gao, and Z. Lv, "Robust EOG-based saccade recognition using multi-channel blind source deconvolution," *Biomedical Engineering/Biomedizinische Technik*, vol. 64, no. 3, pp. 309–324, 2019.
- [7] M. P. Atre and S. D. Apte, "Generalized modeling of body response of acoustic guitar for all frets using a response for single fret," *Applied Acoustics*, vol. 145, pp. 439–444, 2019.

- [8] B. B. Mehta, S. Coppo, D. F. McGivney et al., “Magnetic resonance fingerprinting: a technical review,” *Magnetic Resonance in Medicine*, vol. 81, no. 1, pp. 25–46, 2019.
- [9] M. Ambrosiano, M. T. Bevacqua, T. Isernia, and V. Pascazio, “The tomographic approach to ground-penetrating radar for underground exploration and monitoring: a more user-friendly and unconventional method for subsurface investigation,” *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 62–73, 2019.
- [10] K. Yatabe, Y. Masuyama, T. Kusano, and Y. Oikawa, “Representation of complex spectrogram via phase conversion,” *Acoustical Science and Technology*, vol. 40, no. 3, pp. 170–177, 2019.
- [11] J. Zhuo, L. Ye, F. Han, L. Xiong, and Q. H. Liu, “Multiparametric electromagnetic inversion of 3-D biaxial anisotropic objects embedded in layered uniaxial media using VBIM enhanced by structural consistency constraint,” *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 6, pp. 4774–4785, 2020.
- [12] C. He, A. Cao, J. Chen et al., “Direction finding by time-modulated linear array,” *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 7, pp. 3642–3652, 2018.
- [13] T. Long, Z. Liang, and Q. Liu, “Advanced technology of high-resolution radar: target detection, tracking, imaging, and recognition,” *SCIENCE CHINA Information Sciences*, vol. 62, no. 4, pp. 10–26, 2019.
- [14] D. Jiang, D. Jin, J. Zhuang, D. Tan, D. Chen, and Y. Liang, “A computational model of emotion based on audio-visual stimuli understanding and personalized regulation with concurrency,” *Concurrency and Computation: Practice and Experience*, vol. 33, no. 17, article e6269, 2021.
- [15] S. Sun, B. J. Kooij, A. G. Yarovoy, and T. Jin, “A linear method for shape reconstruction based on the generalized multiple measurement vectors model,” *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 4, pp. 2016–2025, 2018.
- [16] P. Klimek, R. Kreuzbauer, and S. Thurner, “Fashion and art cycles are driven by counter-dominance signals of elite competition: quantitative evidence from music styles,” *Journal of the Royal Society Interface*, vol. 16, no. 151, p. 20180731, 2019.
- [17] P. Lei, M. Chen, and J. Wang, “Speech enhancement for in-vehicle voice control systems using wavelet analysis and blind source separation,” *IET Intelligent Transport Systems*, vol. 13, no. 4, pp. 693–702, 2019.
- [18] A. Massa, G. Oliveri, M. Salucci, N. Anselmi, and P. Rocca, “Learning-by-examples techniques as applied to electromagnetics,” *Journal of Electromagnetic Waves and Applications*, vol. 32, no. 4, pp. 516–541, 2018.
- [19] M. Darbas and S. Lohrengel, “Review on mathematical modeling of electroencephalography (EEG),” *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 121, no. 1, pp. 3–39, 2019.
- [20] A. T. Bukkapatnam, P. Depalle, and M. M. Wanderley, “Defining a vibrotactile toolkit for digital musical instruments: characterizing voice coil actuators, effects of loading, and equalization of the frequency response,” *Journal on Multimodal User Interfaces*, vol. 14, no. 3, pp. 285–301, 2020.
- [21] M. Müller, “An educational guide through the FMP notebooks for teaching and learning fundamentals of music processing,” *Signals*, vol. 2, no. 2, pp. 245–285, 2021.
- [22] M. C. Knaus, “A double machine learning approach to estimate the effects of musical practice on student’s skills,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 184, no. 1, pp. 282–300, 2021.
- [23] K. Liu, “Fast imaging of sources and scatterers in a stratified ocean waveguide,” *SIAM Journal on Imaging Sciences*, vol. 14, no. 1, pp. 224–245, 2021.
- [24] Z. Jamil, A. Jamil, and M. Majid, “Artifact removal from EEG signals recorded in non-restricted environment,” *Biocybernetics and Biomedical Engineering*, vol. 41, no. 2, pp. 503–515, 2021.