

## Research Article

# Optimizing the Design of a Vocal Teaching Platform Based on Big Data Feature Analysis of the Audio Spectrum

Jianhong Hao 

*The Academy of Music, Weifang College, Weifang City, Shandong Province, China 261061*

Correspondence should be addressed to Jianhong Hao; [14211020437@stu.cpu.edu.cn](mailto:14211020437@stu.cpu.edu.cn)

Received 24 March 2022; Revised 14 April 2022; Accepted 22 April 2022; Published 9 May 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Jianhong Hao. 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.

With the development of electronics and communication technology, digital audio processing technologies such as digital audio broadcasting and multimedia communication have been widely used in society, and their influence on people's lives has become increasingly profound. At present, the real-time and accuracy of musical instrument tuners on the market need to be improved, which hinders the design of vocal music teaching system. Based on the BP neural network algorithm and fast Fourier transform algorithm in FPGA, this paper designs a real-time and efficient audio spectrum analysis system, which realizes the spectrum analysis function of music signal. The methods to calculate fast discrete Fourier transform are the FFT algorithm based on time extraction and the FFT algorithm based on frequency extraction. The characteristic of BP neural network algorithm is that it can not only obtain the corresponding estimation results by forward propagation of the input data but also carry out back propagation from the output layer according to the error between the estimation results and the actual results, so as to optimize the connection weight between each layer. This paper proposes to add Nios II system to FPGA processor and adopt cyclone IV in the hardware design of the system, which can be better compatible with the system designed in this paper. In the software part, WM8731 is used to process the audio data. WM8731 consumes very little power to the system, which will effectively improve the processing efficiency of the system. Compared with the original system, the data model obtained after screening and processing of the system model designed in this paper has an algorithm accuracy of more than 90%, among which the audio spectrum clarity of vocal music can reach 95%. Based on the above, the circuit of each module is tested, and in the specific experimental process, the audio frequency spectrum under different conditions is analyzed and data processed. The system can complete the collection and analysis of various music signals in real time, overcome the limitation of single function of traditional tuner, improve the utilization rate of tuner and the clarity of timbre, and also tune a variety of musical instruments and greatly improve the intonation of musical instruments and the utilization rate of tuner, which has a certain practical value.

## 1. Introduction

Due to the rapid development of electronic technology and communication technology, digital audio processing technologies such as digital audio broadcasting and multimedia communication have been widely used in society [1]. Among them, the audio signal needs to be digitized by the tuner before it can be converted into digital signal, then restored to analog signal, and finally converted into sound [2]. The process of vocal music teaching is essentially the process of using scientific means to understand and understand their own “musical instrument”-vocal organ, and

make reasonable adjustment and “transformation,” so that they can finally freely use their “playing” vocal music [3]. At present, most tuners on the market use the principle of vibration, and some high-end tuners use both the principle of sound and the principle of vibration for tuning [4]. However, we generally use tuning software on mobile phones. Although it is convenient, its main disadvantage is that tuning in noisy environment will be inaccurate, because mobile phone software cannot tell whether it is the sound of strings or other sounds, and speaking during tuning will also affect its performance [5]. The application of audio frequency spectrum analysis has the advantages and

characteristics in visualizing the teaching content and mobilizing students' learning enthusiasm, which can not be compared with the traditional vocal music teaching mode.

Therefore, this paper proposes a method that can be applied to other musical instrument tuners, using FPGA technology to realize the analysis and display of the audio spectrum, which can better analyze the audio of the musical instrument, and has reached the optimization of the current vocal music platform. [6]. The principle of the tuner is to transform the input audio signal with fast Fourier transform (FFT), find the frequency of the maximum energy in the frequency range, and then convert the frequency into tone. The vocal music teaching platform system designed and studied in this paper can be applied to vocal music teaching, so that users can realize a comprehensive vocal music learning platform by more clearly displaying and expressing the characteristics of audio spectrum, put learning in the vocal music learning environment, change the previous situation that hinders vocal music teaching due to fuzzy analysis of audio spectrum, and increase the learning effect [7]. This system can be applied to other professional learning systems. The functional model and dynamic model of the teaching and learning system studied in this paper can be applied and promoted, which improves the user's learning interest and increases the learning effect.

The following are the optimization and innovation of the above problems in this paper:

- (1) This paper proposes to design and implement a real-time and efficient audio spectrum analysis system based on the FPGA BP neural network algorithm and fast Fourier transform algorithm. Because the basic composition of the vocal music teaching platform system is based on its hardware and because of the parallel execution of the rules, the operation efficiency can be greatly improved
- (2) Based on the previous vocal music teaching platform, this paper further studies the optimization of audio spectrum analysis and recognition. On this basis, equipment such as tone regulator and loss regulator is embedded in the system [8]. In this way, the system can be more compatible in different teaching environments, which is convenient for vocal music teaching

The paper is mainly divided into the following five parts, and its specific structure is as follows:

The first chapter is the introduction, which describes the research background and significance, and expounds the innovation of the article. The second chapter summarizes the relevant research results from the existing literature at home and abroad and provides the research ideas of this paper [9]. The third chapter is the method part. By introducing the FPGA core processor and related software, the functional connection between the modules is realized, and the feature analysis of the audio spectrum can be more perfect and reasonable with the support of the optimization algorithm. Based on the software design, the core processor pro-

posed in this paper is constructed and equipped with Nios II processor. The fourth chapter is the experiment and data processing part [10]. Test and simulate the circuit board of the hardware, and analyze and organize the test results under the test of the FFT module and the test script. The fifth chapter of this paper is the summary part, which summarizes the optimization of the design proposed in this paper on the original system and the functional shortcomings in some cases.

## 2. Related Work

Armstrong [11] realized that a deep and thorough understanding of the morphological and physiological characteristics of sound movement has become an important prerequisite and consideration for understanding and mastering modern composition technology. Kim et al. [12] put forward a new short video communication mode of vlog in recent years. Its content takes the real life of a vlogger as the theme, such as their daily life, travel diary, and product evaluation. These extremely real fragment life contents have a strong sense of scene, meet the current audience's psychology of needing company, and have a strong resonance. The research of Torija and Flinder [13] shows that spectral analysis can not only play a role in the "display" of timbre in music but also can tune it, so that the rhythm of music can be better improved. Kai et al. [14] propose a real-time software application "Snail" that enables the visualization of sounds and music for tuning instruments and working on pitch intonation. In the 1980s, Sladen and Ricketts [15] a proposed hidden Markov model for the training of acoustic model in speech recognition. At the beginning of the 21st century, Prasad and Yegnanarayana B [16] proposed a bidirectional LSTM network model based on audio signal frame, which is used to classify phonemes, that is, the establishment of acoustic model. Paul's [17] research shows that using energy, zero-crossing rate, fundamental frequency, and spectral peak trajectory as features and using rule-based method, aiming at the segmentation and classification of audio parts extracted from movies and TV programs, it can be divided into voice, music, songs, voice and background music, ambient sound and background music, and silence in real time, achieving 90% classification effect. Xu et al. [18] proposed an unsupervised iterative spectral clustering method and a context-dependent factor scaling method to segment audio into audio elements. Using text retrieval, they proposed a variety of methods for judging the importance of audio elements. For content-based audio retrieval, Liu et al. [19] used the distribution of zero crossing rate as the feature for the voice and music in FM broadcasting and compared the threshold value when judging, which achieved a recognition rate of 98% and strong robustness to the channel. Wen et al. [20] used 13 kinds of audio feature combinations in time domain, frequency domain, and cepstrum domain and various classifiers to classify audio signals from various sources and achieved a false recognition rate of 1.4%. Zhang et al. [21] used the hidden Markov model as the classifier; used unsupervised clustering method to initialize the model; divided the conference recording into silence,

voice, laughter, and others; and then divided the voice into different speakers. Zhang and He [22] proposed a content-based audio classification and retrieval method. They did not use fundamental frequency or spectral distribution as features but counted the distribution characteristics of the parameters of a supervised tree vector quantizer, as a data-driven audio similarity measure. Chen et al. [23] proposed the use of my SQL free database. The overall function of the system is generally divided into the front-end learning system for students. In the back, the system administrator or teacher is responsible for the maintenance of courseware resources. In the front end, JS code is generally used to control the playback of courseware resources, and the data is read at the server. The research by Liu et al. [24] shows that most educational institutions and universities in China have invested a lot of hardware and software resources in online education, one of which is obvious. Most universities in China have set up distance education systems to provide a full set of professional learning for some online learners. Learners take exams online and finally finish their studies, and the schools issue corresponding learning certificates or certificates.

Based on the research of the above-mentioned related work, this paper discusses the connotation and principle of audio frequency spectrum feature analysis and improves and optimizes the hardware facilities of the original vocal music teaching platform system, using the FPGA processor and the Avalon bus technology and BP in the FPGA. Neural network algorithms and fast Fourier transform algorithms provide efficient performance for the functional connection of each software module.

### 3. Methodology

*3.1. Basic Concepts of Audio Spectrum Analysis and Related Theories.* The classification of signals is mainly based on the characteristics of the signal waveform. The audio signal is a one-dimensional signal that changes with time, and the frequency range occupied can reach more than 10 kHz, but the component that has a significant impact on speech intelligibility and intelligibility has a maximum frequency of about 5.7 kHz. According to different characteristics, times can be divided into many categories. From the description of signals, it can be divided into deterministic signals and non-deterministic signals. From the analysis field, it is divided into time domain and frequency domain signals. From the form of signal waveform, it can be divided into continuous time signal and discrete time signal. The basic classification structure of its signal is shown in Figure 1.

In order to study the frequency structure of the signal and the relationship between the amplitude and phase of all frequency components, we should analyze the frequency spectrum of the signal, transform the time domain description of the signal into the frequency domain description with appropriate methods, and express the signal with frequency independent variables. The relationship between time, frequency, and amplitude is shown in Figure 2.

From the mathematical point of view, the power spectrum of a stable stochastic process  $x_n$  is related to the corre-

lation sequence through discrete-time Fourier transform formula. From the perspective of normalized angular frequency, there is the following formula:

$$S_{xx}(\omega) = \sum_{m=-\infty}^{\infty} R_{xx}(m)e^{-j\omega m}, \quad (1)$$

$$S_{xx}(\omega) = |X(\omega)|^2, \quad X(\omega) = \lim_{N \rightarrow \infty} (1/\sqrt{N}) \sum_{n=-N/2}^{N/2} x_n e^{j\omega n},$$

and  $-\pi < \omega \leq \pi$ . The correlation sequence can be obtained by IDFT transform from the power spectrum:

$$R_{xx}(m) = \int_{-\pi}^{\pi} \frac{S_{xx}(\omega)e^{j\omega m}}{2\pi} d\omega = \int_{f_s/2}^{f_s/2} \frac{S_{xx}(f)}{f_s} df. \quad (2)$$

The average power of  $x_n$  sequence over the entire Nyquist sampling interval can be expressed as

$$R_{xx}(0) = \int_{-\pi}^{\pi} \frac{S_{xx}(\omega)}{2\pi} d\omega = \int_{-f_s/2}^{f_s/2} \frac{S_{xx}(f)}{f_s} df. \quad (3)$$

The average power in the frequency band  $[\omega_1, \omega_2]$  and  $0 \leq \omega_1 < \omega_2 \leq \pi$  of a signal can be calculated by integrating the power spectral density in the frequency band. Based on the analysis of the above formula, we can know that signal spectrum analysis is a method to convert time-domain signals into frequency-domain signals. The purpose of spectral analysis is to decompose a complex time-history waveform into several single harmonic components through Fourier transform, thereby obtaining the frequency structure and information of harmonics and phases. Although the Fourier transform method can decompose the signal into a combination of different frequency components, so as to connect the time-domain and frequency-domain characteristics of the signal, the Fourier transform uses a global transformation and cannot indicate the signal's characteristics: time-frequency local properties.

The BP neural network algorithm and fast Fourier transform algorithm are a general method to calculate fast Fourier transform efficiently. The methods to calculate fast discrete Fourier transform are FFT algorithm based on time extraction and FFT algorithm based on frequency extraction. The former divides the time-domain signal sequence according to the even number, while the latter divides the frequency-domain signal sequence according to the odd number. They all rely on two characteristics as follows: one is periodicity; the second is symmetry.

Nios II is a soft core processor. A soft core refers to an IP core that is not configured on the silicon chip and needs to be configured and downloaded to the programmable chip through EDA software. The biggest feature of soft core is that it can be configured by users according to their needs. Nios II soft core processor is a 32-bit soft core processor, which includes three kinds of cores: fast and economical Nios II/e and standard Nios II/s, each of which has its own advantages and disadvantages. Among them, the fast core has the highest performance, the economical core can make the least use of logic resources, and the performance and size

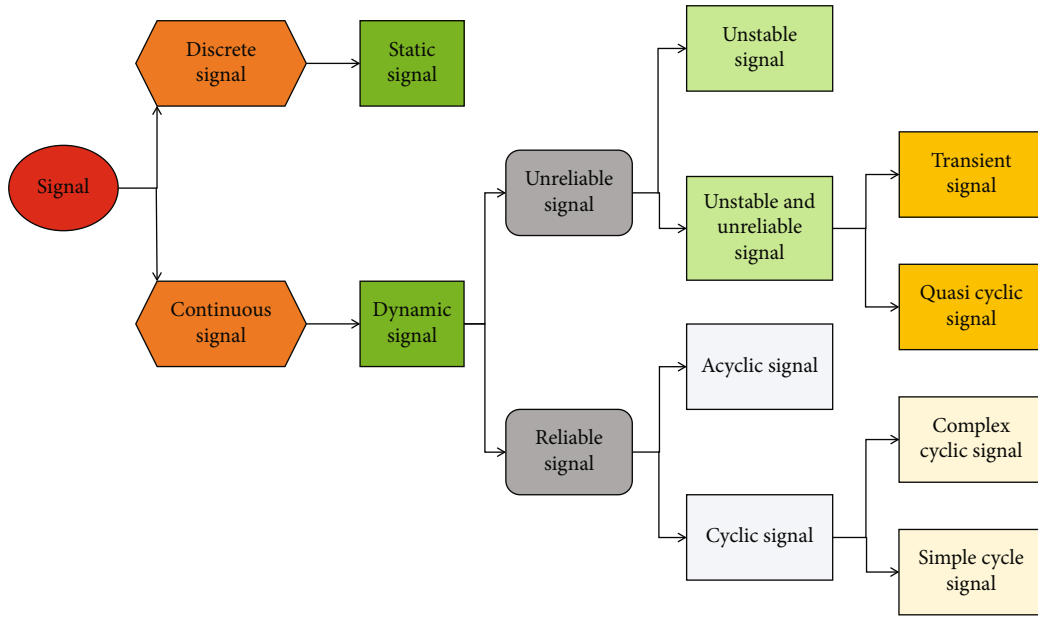


FIGURE 1: Basic classification diagram of signals.

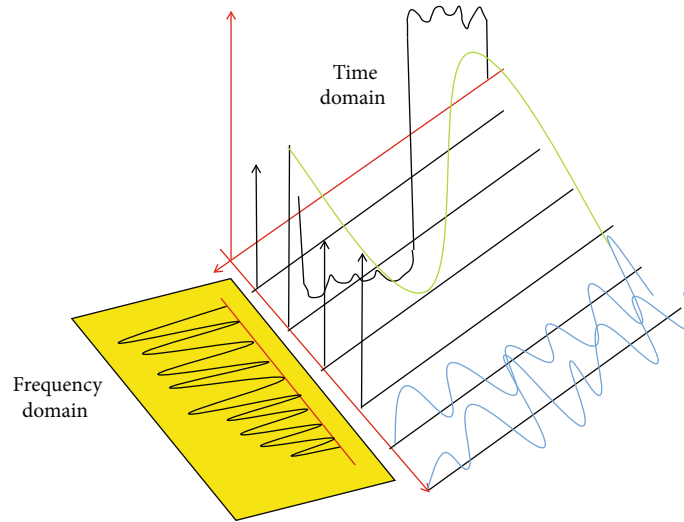


FIGURE 2: The relationship between time, amplitude, and frequency.

of the standard core are superior to the other two cores and run faster. The general structure diagram of NIOS II processor is shown in Figure 3.

From the above figure, we can know the data processing flow of Nios II processor. Firstly, the data processing part is processed in the algorithm logic unit module. Because there is no auxiliary processor in the existing Nios II kernel, the interrupt processor is embedded to improve the overall performance of the system. Nios II's register file includes 32 general-purpose registers and 6 control registers. Nios II structure allows the addition of floating-point registers in the future.

3.2. Design of Audio Acquisition and System Module. The spectrum display system is realized based on Nios II system,

including ADC module, LCD module, RAM module, and Nios II system module. Nios II is the core, which forms a complete main control system through the core processor and peripheral circuits. It can be seen from the overall design block diagram that after the multiple number acquisition module collects the audio signal, it performs  $a/D$  conversion, and the converted data is still stored in the FIFO. Then, the data is output to the FFT calibration block in the FPGA for FFT processing, and then, the data is transmitted with the LCD module through the core processor to transmit the spectrum data of the audio signal after FFT processing to the LCD block, and finally through the analysis and processing of the LCD controller, the audio spectrum is displayed.

Cyclone IV is used as the core processor of the hardware part of the system, which has high capacity and can be well

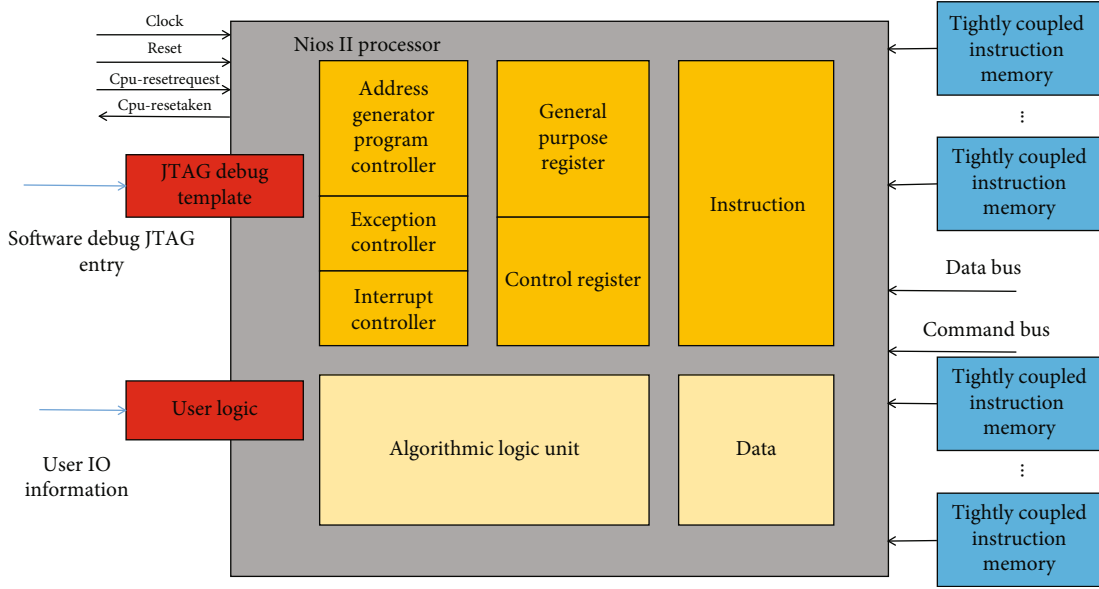


FIGURE 3: General structure diagram of Nios II processor.

applied to program design, so that system developers can reduce the cost and meet the increasing bandwidth requirements. Cyclone IV series devices are designed and used on the basis of low power consumption. There are two types of chips in this series: Cyclone IV E and Cyclone IV GX. In the setting of FPGA module, JTAG configuration and JIC configuration are adopted.

In addition, in the software configuration of the system, because this module is the function of collecting input audio signals and converting digital signals and it is also the core of analyzing the whole audio spectrum characteristics, this paper uses WM8731 to process the above data. The wm8731 is a very low-power, high-quality audio codec and integrated headphone driver designed for portable digital audio applications. Audio signal itself is a typical nonstationary signal. However, in the process of processing, we can assume that it is short-time stationary, that is, in the time period of 10~30 ms, its spectral characteristics and some physical characteristic parameters can be approximately regarded as unchanged. In this way, we can deal with it according to the analysis and processing method of stationary process.

Finally, the core processor needs to be constructed. Altera IP cores are logic function blocks optimized for the FPGA chip structure. Using IP core instead of user-designed logic reduces the development time of chip design and adds new effective logic implementation on the original basis. IP core has the advantages of high efficiency and low cost. Among them, feature extraction, feature data processing, and model design are the focus of this system. The system uses Python as the main development language, because Python supports the mainstream machine learning framework scikit learn and deep learning framework tensorflow and supports the audio feature processing library required by the system, so the system uses Python for feature extraction, feature data processing and model development.

### 3.3. Error in Data Analysis by BP Neural Network Algorithm.

The characteristic of BP neural network algorithm is that it can not only obtain the corresponding estimation results by forward propagation of input data but also optimize the connection weights between layers by backward propagation from the output layer according to the error between the estimation results and the actual results. Adjust the connection weight between each layer according to the error value to gradually reduce the error of the model. Through the repeated training of the model, the model parameters that can minimize the output error can be obtained, and the construction of the model can be completed. Based on the error analysis of the weight of the three-layer BP neural network algorithm model, the following formula and calculation can be obtained:

The output value  $out_{\eta}$  of BP neural network model corresponds to a true value  $out_{\eta}^{\wedge}$ , respectively, so the error function needs to be defined as follows:

$$EL = \frac{1}{2} \sum_{\eta=1}^{\eta'} e^2_{\eta} = \frac{1}{2} \sum_{\eta=1}^{\eta'} \left( out_{\eta} - out_{\eta}^{\wedge} \right)^2. \quad (4)$$

The BP neural network has the property that the random gradient drops when it is used, so based on the above error function, the connection weight is modified reversely, and for each training sample, the connection weight changes to the negative gradient direction, so as to minimize EL.

The specific derivation of the weight is as follows:

Differentiate the error function EL and the link weight  $\lambda_{\alpha\beta}^1$ :

$$\frac{\partial EL}{\partial \lambda_{\alpha\beta}^1} = \frac{\partial EL}{\partial s_{\beta}^1} \cdot \frac{\partial s_{\beta}^1}{\partial \lambda_{\alpha\beta}^1}, \quad (5)$$

$$\frac{\partial EL}{\partial \lambda_{\alpha\beta}^1} = \text{in}_\alpha \cdot \frac{\partial EL}{\partial s_\beta^1}. \quad (6)$$

Calculate  $\partial EL/\partial s_\beta^1$  in (6):

$$\frac{\partial EL}{\partial s_\beta^1} = \frac{\partial EL}{\partial s_\beta^2} \cdot \frac{\partial s_\eta^2}{\partial s_\beta^1}. \quad (7)$$

Because of

$$\begin{aligned} s_\eta^2 &= \sum_{\beta=1}^{\beta'} \lambda_{\beta\eta}^2 \cdot \theta(s_\beta^1) \Rightarrow \frac{\partial s_\eta^2}{\partial s_\beta^1} = \frac{\partial s_\eta^2}{\partial \theta(s_\beta^1)} \cdot \frac{\partial \theta(s_\beta^1)}{\partial s_\beta^1} \\ &= \lambda_{\beta\eta}^2 \cdot \theta'(s_\beta^1), \end{aligned} \quad (8)$$

it can be calculated

$$\frac{\partial EL}{\partial s_\beta^1} = \frac{\partial EL}{\partial s_\beta^2} \cdot \lambda_{\beta\eta}^2 \cdot \theta'(s_\beta^1). \quad (9)$$

Simplify the calculation; make  $\delta_\beta^1 = \partial EL/\partial s_\beta^1$ ; then,

$$\delta_\beta^1 = \frac{\partial EL}{\partial s_\beta^1} = \delta_\beta^2 \cdot \delta_{\beta\eta}^2 \cdot \theta'(s_\beta^1) = \left( \text{out}_\eta - \hat{\text{out}}_\eta \right) \cdot \frac{\partial \text{out}_\eta}{\partial s_\eta^2}. \quad (10)$$

Because of  $\text{out}_\eta = \theta(s_\eta^2)$ , so

$$\delta_\eta^2 = e_\eta \cdot \theta'(s_\eta^2). \quad (11)$$

On this basis, we get

$$\frac{\partial EL}{\partial \lambda_{\alpha\beta}^1} = \text{in}_\alpha \cdot \delta_\beta^1, \quad (12)$$

$$\frac{\partial EL}{\partial \lambda_{\beta\eta}^2} = \frac{\partial EL}{\partial s_\eta^2} \cdot \frac{\partial s_\eta^2}{\partial \lambda_{\beta\eta}^2} = \delta_\eta^2 \cdot \theta(s_\eta^2) \cdot \theta(s_\beta^1). \quad (13)$$

Based on the derivation of the weight gradient above, the weight gradient between the hidden layer and the output layer of BP neural network is affected by the error  $e_\eta$ , the derivative  $\theta'(s_\eta^2)$  of the activation function, and the activation function  $\theta'(s_\beta^1)$ , which means that when the distance between the output layer result  $\text{out}_\eta$  and the actual value  $\hat{\text{out}}_\eta$  becomes smaller, the weight gradient also decreases until the minimum value (or possibly the minimum value) of the error function EL is obtained and the iteration stops. It is not difficult to see that the BP neural network algorithm has an obvious defect; that is, it is easy to fall into local minimum and form local optimum. Because the BP algorithm adopts a gradient descent method, which can make the network weights converge finally, but it cannot guarantee that

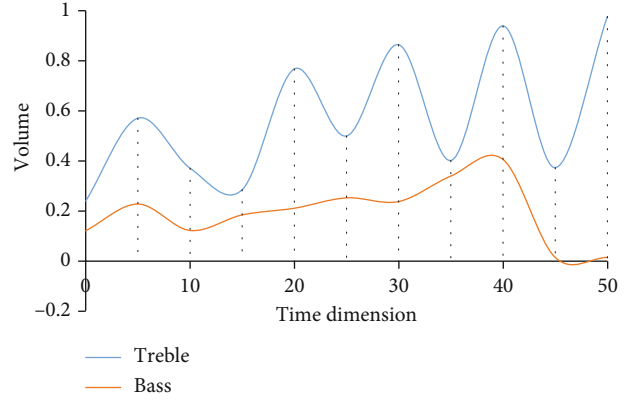


FIGURE 4: Comparison of high and low volume of vocal signal.

the optimal solution is the global optimal solution but may be a local optimal solution.

#### 4. Result Analysis and Discussion

Complete the hardware circuit test and program design module test, and solve the problems found in them, and then test the function of the whole system. The test idea of the whole system is to input music signals with different rhythms and musical instrument audio signals with different tones and observe the change of spectrum diagram with different tones. In this experiment, the function that the spectrogram can change with the tone of the music is first verified, so the vocal signal under different conditions is deliberately input, and the result is shown in Figure 4:

It is found that the spectrogram will change with the pitch of vocal music signal. When the tone of the input vocal music signal suddenly becomes higher, the spectrogram will also suddenly become higher. When the pitch is low or the sound is low, the spectrogram has basically no graphics. It can be seen that the system can complete the function that the spectrogram changes with the change of tone. This is a basic function, but due to the fuzziness of the previous vocal music teaching system on this function, the obtained audio signal has not been strong, which affects the teaching effect. The model mentioned in this paper effectively avoids this situation.

Due to the discreteness of the investigated vocals, the timbre and volume of each user's musical instruments are different, and the audio spectral energy characteristics are very sensitive to the volume. Therefore, it is necessary to adopt the method of energy normalization to ensure the consistency of energy characteristics of audio signals with the same semantics at different volume levels, so as to ensure the robustness of the audio frequency spectrum to the volume of musical instruments. The following are two sets of data models for energy normalization experiments on experimental results:

The first experiment is the parameter distribution diagram of the same audio feature vector at different volume; Experiment 2 is the distribution diagram of each parameter after energy normalization of the same audio feature vector at different volume. It can be seen from Figure 5 that after

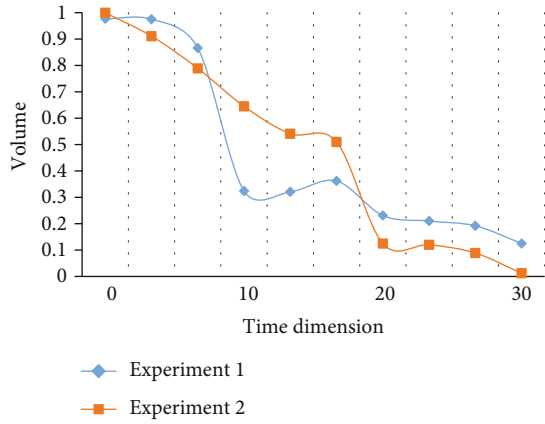


FIGURE 5: Data diagram before and after energy normalization.

energy normalization, the waveforms of increasing and decreasing the volume are consistent with those of the original audio, which shows that the algorithm is robust to the audio spectrum. Figure 5 is the experimental results of frequency band optimization and energy normalization, in which Experiment 1 is the experimental results of frequency band optimization and energy normalization after the signal station enhances bass; Experiment 2 is the experimental result of frequency band optimization and energy normalization to enhance the treble. By taking 30 standard tones, the standard tones are processed by software to enhance the bass and become the enhanced bass test tone, and the standard tones are processed to enhance the treble and become the enhanced high-pitched test tone. It can be seen that after frequency band optimization and energy normalization, the impact of enhanced treble and bass on the recognition results is almost negligible, which shows that the improved algorithm in this paper is more robust to different vocal instruments.

Because we need to get a certain frequency band of vocal music in a certain period of time, we need to detect three different frequency signals generated by the signal source, as shown in Figure 6.

Because the original system can not completely eliminate and filter all frequency signals higher or lower than the cut-off frequency, the processing of audio data is not in place. By processing audio data, the algorithm proposed in this paper can effectively screen and detect all frequency signals above or below the cut-off frequency, without affecting the detection of other frequency signals. The specific experimental data graph is shown in Figure 7.

Obviously, through the screening and filtering of audio data by the system designed in this paper, a clear audio spectrum can be obtained, which greatly improves the vocal music teaching system and improves the teaching sound quality. Due to the high definition of the screened data and graphics, it can also effectively observe the subtle errors of vocal music in teaching and get a good inspection effect.

Set a closed threshold, and the audio packets with correlation less than the threshold are directly judged as out-of-library sounds. In the specific closed value experiment process, three samples of piano sound, harp sound,

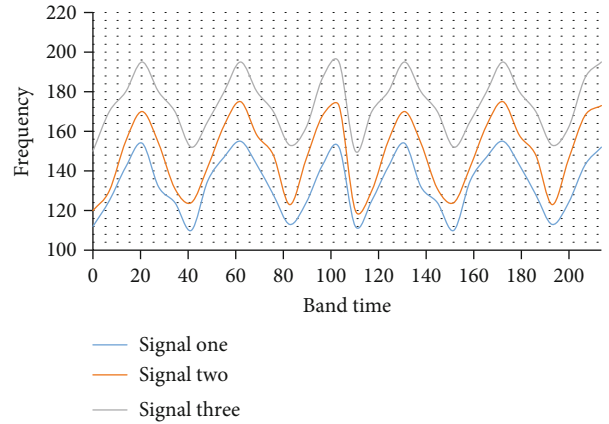


FIGURE 6: Simulation results of signals with different frequencies.

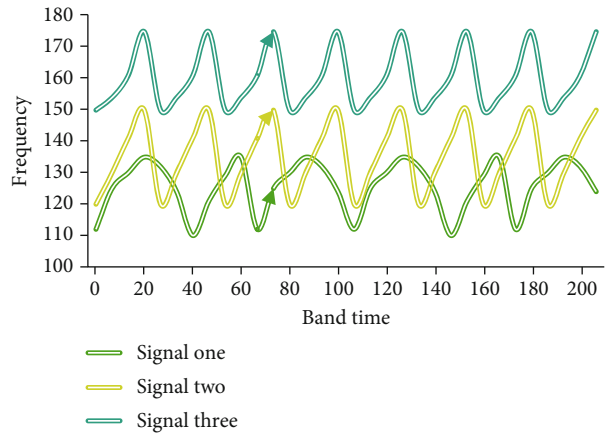


FIGURE 7: Results of signal screening at different frequencies.

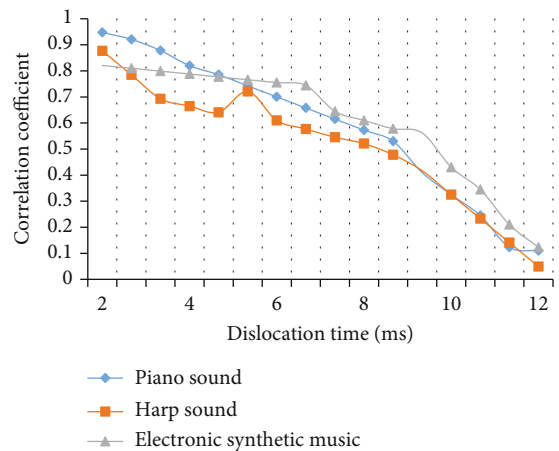


FIGURE 8: The relationship between the dislocation time and the correlation coefficient of the three samples under the condition of a noise ratio of 5 dB.

and electronic synthetic music were selected for testing, and there was no original dislocation in the test. Figure 8 shows that the noise ratio of the three samples of piano

sound, harp sound, and electronic synthetic music is 5 dB. Through the system designed in this paper, the above three samples are tested, and the results shown in Figure 8 are obtained.

The time shift has a great impact on the correlation. Under the condition of ensuring the signal-to-noise ratio, combined with the inevitable time shift in practice, when the shift is about 30 ms, the average correlation coefficient can be above 0.25. In the actual operation process, we should pay attention to denoising and time alignment. The smaller the noise is, the smaller the time dislocation is, and the more accurate the discrimination result is.

After many tests and screening experiments on the actual collected vocal music audio, the closed value is set to 0.3, so that the nonvocal music audio captured by the port can be excluded. At the same time, the number of Euclidean distance calculation files is reduced, and the purpose of reducing the calculation amount of BP neural network algorithm and fast Fourier transform algorithm and increasing the accuracy of recognition results is achieved. Compared with the original vocal music teaching system, the accuracy of system model recognition results proposed in this paper is increased by 85%.

## 5. Conclusions

Based on the BP neural network algorithm and fast Fourier transform algorithm in FPGA, a real-time and efficient audio frequency spectrum analysis system is designed and implemented. This system can complete the identification and analysis of various music signals in real time, can overcome the limitations of traditional tuners, and can tune a variety of musical instruments and screen and detect effective audio spectrums, which greatly improves the intonation of musical instruments and overcomes the limitation that a tuner can only adjust one musical instrument is improved, and the utilization rate of the tuner and the clarity rate of the timbre are improved, which has practical application value. During the experiment, only a simple audio acquisition device was built without any clock calibration and denoising, so the audio effect was poor. However, compared with the original system, the data model obtained after the screening and processing of the system model designed in this paper has an algorithm accuracy of more than 90%, among which the audio spectrum definition of vocal music can reach more than 95%. At the same time, some experiments show that, in actual operation, the model of vocal music teaching system designed in this paper has been tested and screened for many times on the actual collected vocal music audio, reducing the calculation amount of BP neural network algorithm and fast Fourier transform algorithm, which has greatly improved the efficiency of the system, and at the same time, compared with the original vocal music teaching system, the accuracy rate of the system model recognition proposed in this paper has increased by 85%. All these have significantly optimized the original vocal music teaching system and made it more practical for vocal music teaching.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

No competing interests exist concerning this study.

## References

- [1] D. Baskent, J. Libert, D. Vickers, and E. Gaudrain, "The perception of vocal characteristics in normal-hearing and cochlear-implanted children," *Journal of the Acoustical Society of America*, vol. 138, no. 3, pp. 1919–1919, 2015.
- [2] J. Gonzalez and T. Cervera, "The effect of MPEG audio compression on multidimensional set of voice parameters," *Logopedics, Phoniatrics, Vocology*, vol. 26, no. 3, pp. 124–138, 2001.
- [3] H. Relañó-Iborra, T. May, J. Zaar, C. Scheidiger, and T. Dau, "Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 2670–2679, 2016.
- [4] Y. Wan, X. Wang, R. Zhou, and Y. Yan, "Automatic piano music transcription using audio-visual features," *Chinese Journal of Electronics*, vol. 24, no. 3, pp. 596–603, 2015.
- [5] R. D. Patterson, M. Andermann, S. Uppenkamp, and A. Rupp, "Neuromagnetic correlates of the vocal characteristics of vowels in auditory cortex," *Journal of the Acoustical Society of America*, vol. 138, no. 3, pp. 1919–1919, 2015.
- [6] A. Abeysinghe, M. Fard, R. Jazar, F. Zambetta, and J. Davy, "Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise," *The Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 193–201, 2021.
- [7] A. Kolahan, E. Roohi, and M. R. Pendar, "Wavelet analysis and frequency spectrum of cloud cavitation around a sphere," *Ocean Engineering*, vol. 182, no. 15, pp. 235–247, 2019.
- [8] A. Luque, J. Gómez-Bellido, A. Carrasco, and J. Barbancho, "Optimal representation of anuran call spectrum in environmental monitoring systems using wireless sensor networks," *Sensors*, vol. 18, no. 6, p. 1803, 2018.
- [9] T. Shimabukuro and K. Funaki, "Vandermonde transform for complex residual based on complex speech analysis," *Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 3057–3057, 2016.
- [10] Q. Ren, J. Peng, and H. Chen, "Amplitude information-frequency characteristics for multi-frequency excitation of underwater active electrolocation systems," *Bioinspiration & Biomimetics*, vol. 15, no. 1, p. 17, 2020.
- [11] C. M. Armstrong, "These vacuum devices stood guard during the Cold War, advanced particle physics, treated cancer patients, and made the Beatles sound better," *IEEE Spectrum*, vol. 57, no. 11, pp. 30–36, 2020.
- [12] S. Kim, R. Bakker, and M. Ikeda, "Comparing preferred spectral balances of listeners from four countries," *Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 3391–3391, 2016.
- [13] A. J. Torija and I. H. Flindell, "The subjective effect of low frequency content in road traffic noise," *Journal of the Acoustical Society of America*, vol. 137, no. 1, pp. 189–198, 2015.



- [14] K. Yu, R. E. Hudson, Y. D. Zhang, K. Yao, C. Taylor, and Z. Wang, "Low-complexity 2D direction-of-arrival estimation for acoustic sensor arrays," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1791–1795, 2016.
- [15] D. P. Sladen and T. A. Ricketts, "Frequency importance functions in quiet and noise for adults with cochlear implants," *American Journal of Audiology*, vol. 24, no. 4, pp. 477–486, 2015.
- [16] R. S. Prasad and B. Yegnanarayana, "Determination of glottal open regions by exploiting changes in the vocal tract system characteristics," *Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. 666–677, 2016.
- [17] J. D. Paul, "Re-creating the sigsaly quantizer: this 1943 analog-to-digital converter gave the allies an unbreakable scrambler - [resources]," *IEEE Spectrum*, vol. 56, no. 2, pp. 16-17, 2019.
- [18] Z. Xu, C. Ao, and B. Huang, "Channel capacity analysis of the multiple orthogonal sequence spread spectrum watermarking in audio signals," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 20–24, 2016.
- [19] F. Liu, Z. Zhang, R. du, J. Sheng, and C. Huang, "Frequency-angle spectrum hole detection with Taylor expansion based focusing transformation," *IEEE Transactions on Mobile Computing*, vol. 19, no. 10, pp. 2330–2343, 2020.
- [20] X. Wen, Y. Han, J. Liu, J. He, and J. Wang, "Polarization squeezing at the audio frequency band for the rubidium D<sub>1</sub> line," *Optics Express*, vol. 25, no. 17, pp. 20737–20748, 2017.
- [21] Y. Zhang, Z. Xu, and B. Huang, "Channel capacity analysis of the generalized spread spectrum watermarking in audio signals," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 519–523, 2015.
- [22] X. Zhang and Q. He, "Time–frequency audio feature extraction based on tensor representation of sparse coding," *Electronics Letters*, vol. 51, no. 2, pp. 131-132, 2015.
- [23] X. Z. Chen, Y. H. Liu, C. C. Yin et al., "Three-dimensional inversion of controlled-source audio-frequency magnetotelluric data based on unstructured finite-element method," *Applied Geophysics*, vol. 17, no. 3, pp. 349–360, 2020.
- [24] S. Liu, L. Zhao, G. Fang, C. Hu, and Y. Ge, "Nonlinear aerodynamic characteristics and modeling of a quasi-flat plate at torsional vibration: effects of angle of attack and vibration amplitude," *Nonlinear Dynamics*, vol. 107, no. 3, pp. 2027–2051, 2022.