

RESEARCH

Open Access



# Research on vocal sounding based on spectrum image analysis

Jian Sun<sup>1,2</sup>

## Abstract

The improvement of vocal singing technology involves many factors, and it is difficult to achieve the desired effect by human analysis alone. Based on this, this study based on spectrum image analysis uses the base-2 time-selection FFT algorithm as the research algorithm, uses the wavelet transform algorithm as the denoising algorithm, and combines comparative analysis to discuss the mechanism of vocal music, the state of vocalization, and the vocal quality of vocalists in vocal music teaching. Simultaneously, this study compares the singer's frequency, pitch, overtones, harmonics, singer formants, etc., and derives the characteristics of vocal vocalization under different conditions, and can be extended to all music vocal studies. Research shows that this research method has certain practicality and can provide theoretical reference for related research.

**Keywords:** Vocal music, Image analysis, Music spectrum, Singing, Image processing

## 1 Introduction

The research on vocal singing technology has a long history. In the process, many experts and scholars have continuously studied and summarized this. After repeated exploration and continuous practice, some relatively perfect technical methods have been formed. However, through my own review, it is found that the domestic summary of these singing methods is only in the teaching of words, and the intuitive unified description is basically blank, which is quite inconvenient for us to correctly learn, recognize, and evaluate vocal music. Although the current relevant research report has certain results, considering the vocal cord conditions, the cultural background will bring some deviations to the experimental results. Therefore, we doubt whether their data results can represent the level of the subjects in our country. Whether these tests are in line with the cultural background is indeed a difficult problem to be verified, and it is also an urgent need for domestic research to carry out such research. In addition, there are some differences in different kinds of music, and there are few studies related to vocal music in related research. Therefore, this study combines image processing to study the occurrence of vocal music [1].

Wang et al. proposed a normalized model of pitch perception based on the unique resolution of the human ear to different frequency sources. This algorithm has great significance in multi-frequency estimation techniques [2]. Wu et al. simplified the pitch model by using the Bark scale to simulate the human ear frequency response, which greatly reduced the amount of computation. The method calculates the autocorrelation coefficient of the signal through two channels, and the multi-channel signal autocorrelation coefficient in the non-ERB scale, and finally calculates the respective fundamental frequency values one by one by scale stretching and linear interpolation. Similarly, the cycle-based algorithm was used by Staley and Sethares. Similarly, based on the establishment of the pitch model, Cheveigne et al. recursively screened the estimated fundamental frequencies in the poetry selection [3].

Guan et al. used the new technology to develop one of the most advanced automatic music annotations at that time. The multi-frequency estimation method in this system was also in a leading position at that time. They applied computer scene analysis in the system, assuming that the sinusoidal trajectory is a musical note, and in terms of methods, they cite the principle of perceptual aggregation, including harmonicity, frequency, timbre, and melody. Guan et al. designed musical instrument models for the instruments corresponding to each note.

Correspondence: [sunjianchengdu@163.com](mailto:sunjianchengdu@163.com)

<sup>1</sup>Music college, Mahasarakham University, Mahasarakham 44150, Thailand

<sup>2</sup>Huayang High School, Chengdu 610213, China

These models can solve the problem of spectral energy distribution when notes of different frequencies overlap. In addition, they analyze the probability of occurrence of notes in the case of a given chord and use Markov chains to encode and transfer chord probabilities. The method uses a Bayesian network for linear prediction for bottom-up analysis, and top-down processing includes predictive note spectral components and chord prediction [4].

Meng models the short-time spectrum of the music signal. In the model, a certain number of harmonics constitute each tone model, and each harmonic component is modeled by a Gaussian distribution centered on the fundamental frequency integer multiple. Meng designed a multi-agent structure to track the fundamental frequency weights of consecutive frames and iteratively updated each pitch model and its weight with the method of calculating the expectation maximization. The algorithm can successfully track the pitch lines of the music signal and its melody. Although the system is more complicated, the core EM algorithm is relatively easy to implement in the experiment. In addition, the algorithm can correctly estimate the weight of all fundamental frequencies, but the disadvantage is that the primary fundamental frequency can usually only be estimated [5].

Qiu et al. improved on the idea of weighted mixed-tone models. However, they modeled each set of harmonics in the tone model instead of modeling each harmonic individually. They believe that although the amplitude of the notes in the signal will change, the harmonic structure of each note is unique and stable. The harmonic structure here refers to the relative relationship between the harmonics. Based on the relationship between harmonic frequencies, Qiu et al. chose to model a set of harmonics using a constrained mixed Gaussian model. It can be seen in the experiment that the constraint to be used in the system reduces the parameters to be estimated in the system. In this modeling mode, the signal model is established and finally the parameter estimation result is given based on the expectation maximization [6].

Spectrum iterative deletion method was adopted by Claudio. In order to reduce the influence of instruments of different timbres as much as possible, the algorithm firstly pre-whitens the signals and then calculates the sum of the energy of the corresponding harmonic frequencies in the frequency domain for each candidate fundamental frequency. Among them, the person with the greatest energy is considered to be the fundamental frequency. Then, the algorithm subtracts the fundamental frequency and its harmonic energy from the original spectrum according to a certain ratio. At this time, the algorithm estimates the next fundamental frequency,

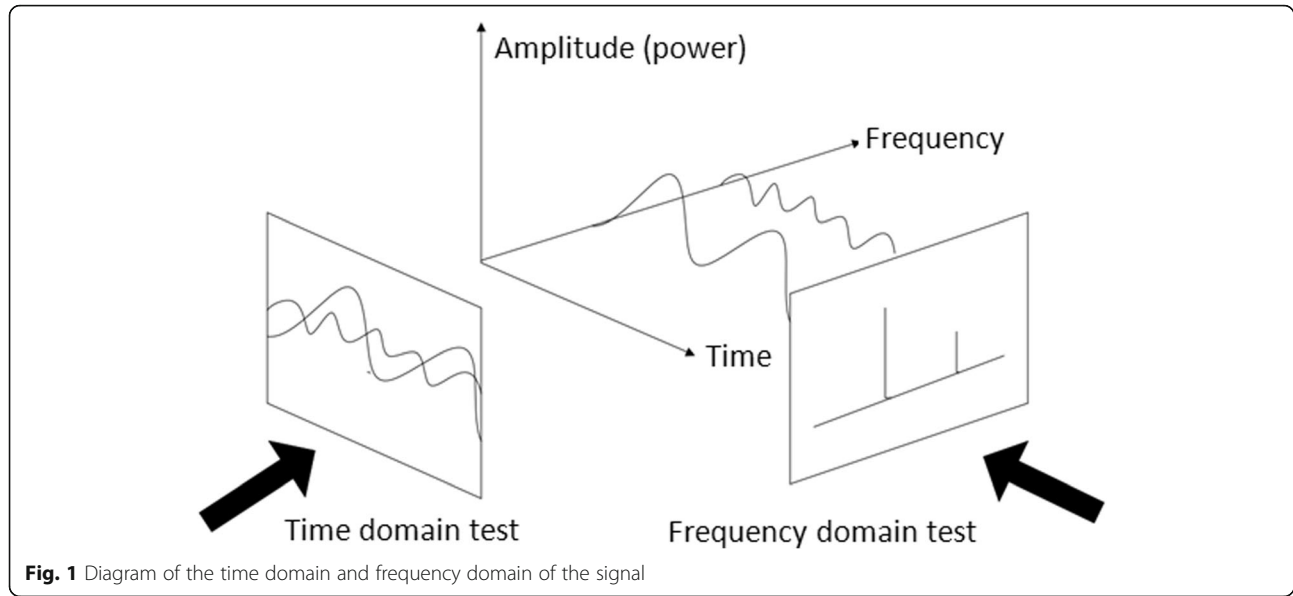
and then iteratively estimates. The algorithm has a lot of technical links that need special attention, such as parameter estimation during pre-whitening, proportional deletion coefficient, and judgment of dominant fundamental frequency estimation. The algorithm tested chords randomly synthesized from a variety of instrumental notes, and the test yielded the desired results [7].

Zabalza sees the fundamental frequency estimation as a multi-label classification problem, which is completely different from the other methods. Each signal frame in the algorithm is considered to be a sample, the fundamental frequency is treated as a label, and the algorithm extracts the features of the spectrum of the signal. They collect all samples of this fundamental frequency for each fundamental frequency, thus training a support vector machine classifier. The algorithm considers the classification result obtained by the classifier as the result of the fundamental frequency estimation. Zabalza et al. chose the piano signal for testing and the results were as expected. However, although the algorithm is novel, it needs to deal with two difficulties: First, although the algorithm uses a one-to-many classifier, the parameter space is greatly reduced, but the fundamental frequency classifier is split, and it also violates the close relationship between the notes. Second, the algorithm needs to train up to dozens of classifiers, which greatly increases the amount of computation [8].

In view of the current understanding of image processing in vocal research topics, the purpose of this study is as follows: First, this article is to facilitate a clear understanding of the current status and trends of research on this topic in the world, as well as some unresolved issues and development limitations, so as to better fill the gap. Second, the author's collation of the adjectives of singing technology can be used as a basic research for other researchers, thus providing an innovative basis for future research.

## 2 Research methods

In this study, spectral image analysis combined with image processing was used for vocal studies and spectral image analysis, that is, the analysis method of the present study is a method of expressing the amplitude and phase of the time domain signal on the frequency coordinate axis and then analyzing the signal in the frequency domain. Figure 1 shows the relationship between the signal time domain and the frequency domain. This paper analyzes the spectrum image of the signal, which can be divided into convolution, correlation, and transformation. Among them, the most basic one is the discrete Fourier transform. Since the amount of calculation directly using the DFT is proportional to the square of the length  $N$  of the operation interval, the amount of calculation is very large. Therefore, by using the



periodicity, symmetry, and other properties of the twiddle factor, the DFT of the long sequence is decomposed into short-order DFTs one by one, and the amount of computation is reduced, which is the fast Fourier transform (FFT) algorithm. The base-2 time-selection FFT algorithm [9] is used in this paper.

By definition, the DNF formula for a finite-length sequence  $x(n)$  of length  $N$  is as shown in Eq. (1):

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk}, k = 0, 1, \dots, N-1 \quad (1)$$

$$W_N^{nk} = e^{-j\frac{2\pi}{N}nk} = \cos\left(\frac{2\pi}{N}nk\right) - j \sin\frac{2\pi}{N}nk \quad (2)$$

In the actual application process, it can be clearly felt that the FFT algorithm is greatly improved compared with the direct calculation of DFT, especially the larger  $N$ , the effect is more obvious. In addition, the computational complexity of the FFT algorithm is significantly less than that of the DFT algorithm, so the FFT algorithm is more suitable for this study.

This study starts with the butterfly operation of spectrum image analysis and optimizes the FFT algorithm to make it suitable for embedded system. On this basis, we propose an audio dual spectrum identification method for the problem of music and color mismatch. Regardless of the type of music, from the perspective of spectral image analysis, their audio signals are composed of various harmonics of the sinusoidal signal, and the spectrum of the audio signal is its essential feature. Therefore, in this experiment, we collected a multi-track MIDI main melody demo of standard music chords. We mainly perform spectral feature extraction on it, and the frequency value of the international standard sound A<sub>4</sub> is

major is 440 Hz. Then, according to the calculation of the 12 average law, the adjacent semitones differ by 2 to 12 times, and the theoretical frequency values of the other majors are obtained [10]. Table 1 shows the frequency values corresponding to the major.

The spectrum of the audio signal in the vocal utterance and the spectrum of which music chords are corresponding or similar are determined by the frequency comparison method. There are many ways to compare. The simplest can be determined by the maximum intensity of the spectrum, i.e., the maximum intensity method of the contrast spectrum. We need to use Bayesian formula to calculate the same or similar probability, which is widely used in scientific research with its profound thoughts.

We set the natural state  $\theta$  to have  $k$  kinds, namely  $\theta_1, \theta_2, \dots, \theta_k$ .  $P(\theta_i)$  represents the prior probability of  $\theta_i$  occurring in the natural state.  $P(x|\theta_i)$  represents the probability that the event is  $x$  in the state  $\theta_i$  condition. The full probability  $P(x)$  is the probability that  $x$  may occur in various states, i.e., [11]

$$P(x) = \sum_{i=1}^k P(x|\theta_i)P(\theta_i) \quad (3)$$

**Table 1** Comparison of the computational quantities of basal-2 FFT and direct computation DFT

	Complex multiplication	Complex addition
Basal-2 FFT	$N/2 \log_2 N$	$N \log_2 N$
Direct computation DFT	$N^2$ numbers	$N(N-1)$

$$P(x|\theta_i) = \frac{P(x|\theta_i)P(x)}{\sum_{i=1}^k P(x|\theta_i)P(\theta_i)} \quad (4)$$

When using the Bayes formula, it is necessary to construct a probability model based on the spectrum of the audio signal. The statistical calculation is used to compare the amplitude of the audio signal sampling spectrum with the peak of the standard music chord spectrum to determine the music chord corresponding to the played audio signal, and then the color can be determined.

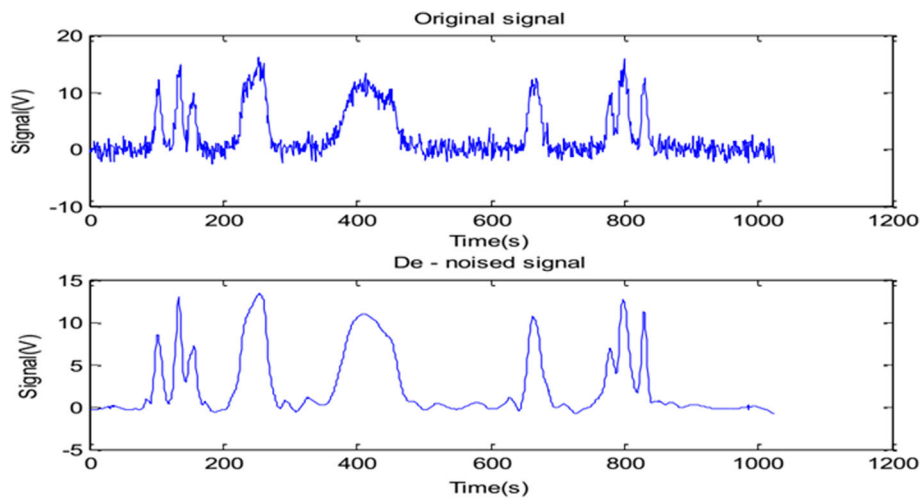
The vocal vocalization will produce noise during the actual extension, and the abovementioned audio format conversion will inevitably generate noise during the conversion process, which will seriously affect the accuracy of the audio dual spectrum identification method. Therefore, this study uses wavelet transform to remove high-frequency noise interference. When the wavelet function  $\Psi(t)$  is a real function, the one-dimensional wavelet continuous transformation formula is as shown in Eq. (5) [12].

$$W_f(a, b) = |a|^{1/2} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \quad (5)$$

Among them,  $a$  is the scale factor and  $b$  is the displacement factor. There are three existing wavelet denoising methods, namely, modulus maximal reconstruction denoising, spatial correlation denoising, and wavelet domain threshold denoising. By comparing the qualitative methods of the three denoising methods, and finally considering the calculation and denoising effects, we choose to use the third wavelet domain threshold denoising. Threshold denoising in the wavelet domain is to select appropriate thresholds on different scales, zero

the wavelet coefficients for Gaussian noise, and preserve the wavelet coefficients larger than the threshold, so that the noise in the audio signal is suppressed, and finally reconstruct the signal to obtain the optimal estimation of the effective signal. Hard threshold denoising or soft threshold denoising should be selected in wavelet domain threshold denoising. By analyzing in MATLAB software, soft threshold denoising is more suitable for our experiments. Meanwhile, the threshold is divided into three thresholds, the Plealty threshold, the Birge-Massart threshold, and the default threshold. Figure 2 is a comparison of the original audio signal with the denoised audio signal. It is obvious that the noise has been processed. In the experiment, we chose to use the Birge-Massart threshold in the soft threshold and select the sym4 wavelet in the wavelet base, that is, we use sym4 wavelet to decompose the audio signal in four layers, and the wavelet base and the number of decomposition layers can be changed, which is equivalent to the wavelet transform formula [13].

The sound of vocal singing is a compound sound, which is composed of the pitch of the singing and other partial sounds. Vocal training is an extremely important part of the vocal teaching process. Both teachers and students need to use the auditory system to judge and identify the sounds produced during singing. These include distinguishing the pitch, strength, and rhythm, as well as a series of phonetic factors that distinguish the physiological state of the singing and the vocabulary pronunciation and language sensation. It can be said that the process of vocal music learning is actually a process of establishing the correct sound concept, and the establishment of this concept depends to a large extent on the sensitivity of the vocal learner's auditory system. The quality of hearing has a very direct relationship with the



**Fig. 2** Comparison between the original audio signal and the denoised audio signal

physiological conditions and experience of the listener. The most obvious thing about the sound spectrum description and identification of the pronunciation of multimedia computer technology is that it has accuracy and intuitiveness. During the vocal training process, we recorded the singer's singing voice through a computer and observed various subtle parameters of the emitted sound through the function chart and the sound wave table displayed on the computer screen. This is a positive complement to the human auditory system in an objective perspective. The use of the computer to participate in the vocal music teaching process has a positive effect on the quality of the vocalist's pronunciation, as well as the analysis of singing and speech intelligibility. The following is a spectrum image analysis test for vocal sounds and the like [14].

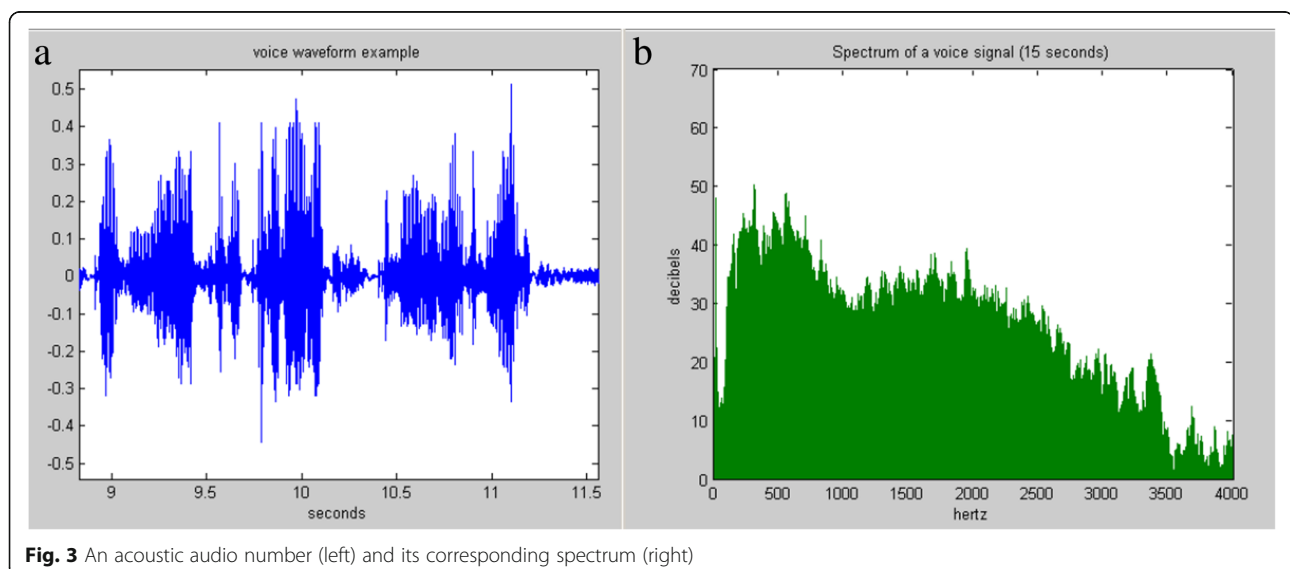
By analyzing the vocalist's pronunciation through spectral image analysis, you can observe the pitch of the singing voice and various overtones, partial sounds, and sounds in different frequency segments. The analysis results from the average of the sound frequency analysis in a certain time zone, which we call the "frequency spectrum." In fact, it is a frequency and sound intensity distribution map of each harmonic column of singing. This frequency spectrum shows the sound intensity profile of the sound of singing at different frequencies, and its sound intensity is expressed in decibels. Usually, we give a more specific "sound form" reference frame from the perspective of musical acoustics. It allows us not only to understand the sound in the abstract concept of the "anapharyngeal cavity, oral cavity, chest cavity, head cavity, high position resonance" pointed out by the abstract general auditory organs, but also to visually see the visual form and specific numerical values. This method of analysis is specific, intuitive, and accurate

[15]. In addition, the graphics and data provided by the computer can not only objectively reflect the characteristics of pitch, sound intensity, etc., but also assist the auditory to further analyze the timbre, resonance, and other factors of sound of singing.

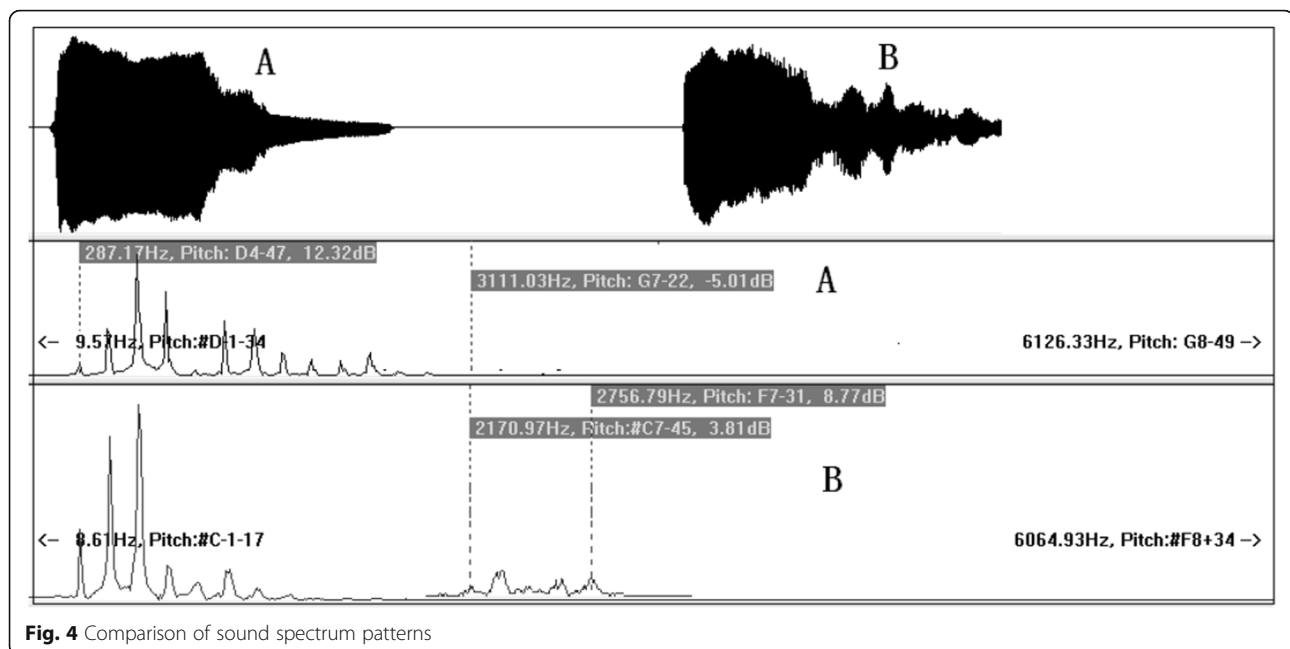
### 3 Results

In the study, a group of normal vocal music was first tested to study the rationality of the operation of the research method. In Fig. 3, we can see an acoustic audio number (Fig. 3 left) and its corresponding spectrum (Fig. 3 right). It can be seen from Fig. 3 that the algorithm of this study is stable and can be used for subsequent research.

It can be seen from the measurement that the pitch of the vocals during singing is in a relatively narrow range, generally between 100 and 1500 Hz. In addition to the high pitch frequency, the singing voice has many overtones and partials. Therefore, the singing voice spectrum test also found that trained singers, in addition to the pitch, will also have a peak of resonance in the overtone range of 2800–3200 Hz, which we call "singer formant" or "Fourth formant." However, without professionally trained singers, there is no obvious singer formant. The "metal sound" and "penetrating sound" often mentioned in vocal music teaching are the auditory feelings of this singer at the formant. The spectrum analysis system of the computer can be used to analyze, compare, and compare the acoustic spectrum images of outstanding singers. The principle of "singing the formant" can be used in the teaching to study and analyze the sound quality of the students singing, so as to implement the corresponding teaching methods. Figure 4 is a comparison of the acoustic spectrum patterns of unseen players and professional singers singing the same pitch. The







method of sound spectrum testing breaks the traditional vocal teaching method that has always been subjective. The test method injects objective and rational scientific factors and further increases the knowledge of singing acoustics and vocal physiology. In Fig. 4, B is a singer who has been trained for a long time, and A is an untrained person.

The soprano vocalist's vocal cords are shorter and thinner, with vocal cord lengths typically ranging from 8 to 11 mm. A short vocal cord is more convenient for high-pitched singing. However, the soprano without rigorous training has a thinner tone in the high-pitched sound zone, and its tone in the middle and bass sound zones sounds hollow and illusory. The excellent, highly trained soprano singer is in a unified state in the tone of each pitch. The characteristics of the sound are bright, crisp, gorgeous, strong, and both round and sweet, and the

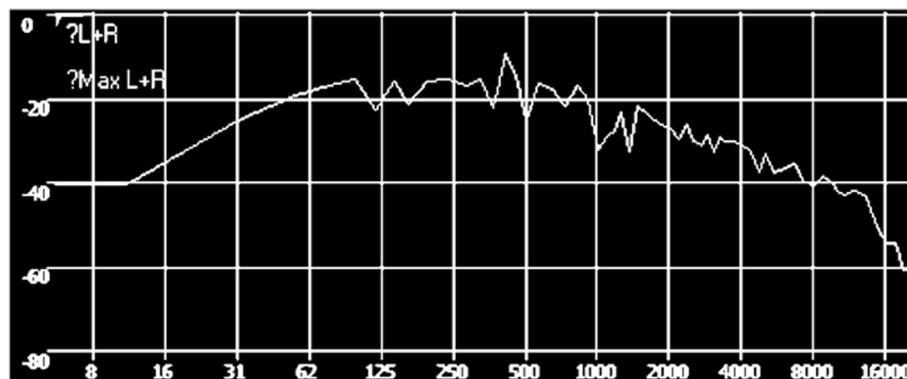
sound is rich and varied. The author's test samples for the analysis of acoustic and acoustic spectrum images are famous soprano singers, Bart and Carava. The first is the spectrum image analysis of the soprano singing. As shown in Fig. 5, this is the d sound of Bart's vocal group of [a] vowels, which has a frequency of about 580 Hz.

As shown in Fig. 6, this is the d sound image of small group two that Carava [a] uses vowel singing, which has a frequency of about 580 Hz.

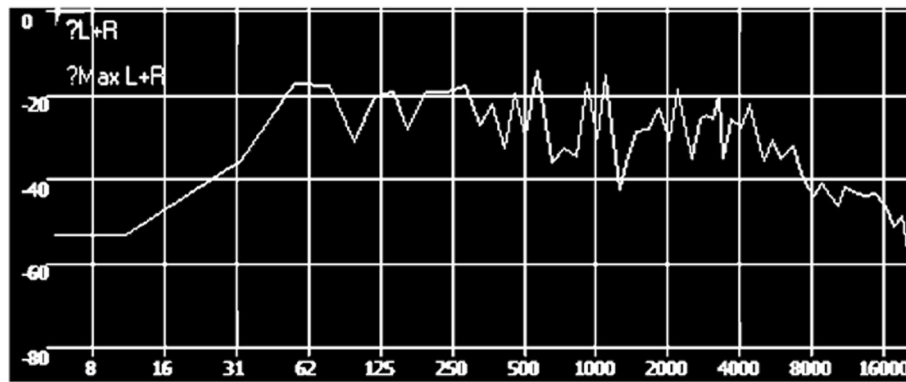
As shown in Fig. 7, this is the g sound image of small group two that Bart uses [o] vowel singing, which has a frequency of about 780 Hz.

Figure 8 shows carawa singing a sound of the small group two with the [o] vowel, which has a frequency of about 880 Hz.

The voice of the mezzo-soprano is slightly longer than the soprano, and its length is between 11 and 14 mm.



**Fig. 5** The d sound image of small group two that Bart [a] uses vowel singing



**Fig. 6** The d sound image of small group two that Carava uses [a] vowel singing

The relatively bright soprano is softer, fuller, deeper, and richer. The excellent mezzo-soprano singer has a wide range of sounds. The middle and low sound areas are wide, sweet, and sleek, and the treble sound area is not bright, but the sound is strong. Sound samples for testing: The experimental samples for the analysis of acoustic and acoustic spectrum images in this paper are famous mezzo-soprano singers, Honre and Bortoli.

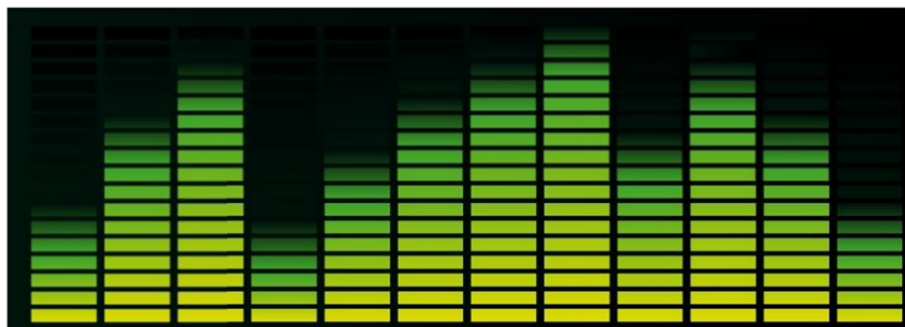
Figure 9 is a spectrum image analysis of mezzo-soprano in singing. As shown in the Fig. 9, Horne sings a sound of the small group with [o] vowel, in which the frequency is about 220 Hz.

As shown in the Fig. 10, Bortoli sings b sound of the small group with [i] vowel, in which the frequency is about 490 Hz.

From the spectrum image analysis chart of Fig. 4, we can see that the spectrum of the untrained person's sound spectrum gradually decays after the peak of the pitch frequency, and we can clearly see that there is almost no peak in the high frequency in Fig. 4 (A). However, professional singers show higher formants in the frequency range of 2170 to 2756 Hz.

The method of acoustic spectrum testing breaks the traditional vocal teaching method that people have always been subjective. The method of sound spectrum

testing injects objective and rational scientific factors into the teaching of vocal music and further enhances the knowledge of singing acoustics and vocal physiology from the simple physiological experience of the past. At the same time, this method has changed the bad learning habits of students who only pay attention to practice and despise principle theory. Vocal music teaching can not only cultivate singers, but should pay equal attention to vocal theory research, vocal practice singing, and vocal teaching methods. In addition, this method can fully reflect the comprehensive ability of vocal singing talents. The use of computer multimedia for spectrum image analysis not only provides a method for identifying and analyzing voice timbre for vocal training and teaching, but also stores and transmits the measured sound spectrum as a data file and also provides a more convenient management method for scientific research and teaching record data and organization. The use of this method can be used for singers, vocal learners (different degrees), or singers of different singing styles, and can make detailed and long-term data analysis and comparison. The analysis results have a more objective reference for the identification of the vocalist's voice, the adjustment of the singing during the singer's learning process, and the selection of talents.



**Fig. 7** The g sound image of small group two that Bart uses [o] vowel singing



**Fig. 8** The g sound image of small group two that Carava uses [o] vowel singing

#### 4 Discussion

In the spectrograms of Figs. 5 and 6, both soprano singers have a high degree of spectral consistency. First, there is no sharp spike-like map of its pitch and overtone. Through this relatively broad pitch peak, it is known that the application of vibrato during singing is reasonable. The amplitude is between a small second, and the frequency of change is about six times per second. Corresponding to such an auditory feeling is that the sound thickness is symmetrical and round without a thin sharpness. Both singers' resonance peaks are around 3000 Hz and the sound intensity of the formant is high, and the sound intensity of the pitch is almost the same. It can be seen from the Fig. that the resonance energy of the formant has a wider "base" at 3000 Hz, which produces vocal transparency in the human ear but does not lose the heavy hearing experience. At the same time, its first overtone and second overtone are higher in intensity than other overtones. We know that the first overtone and the second overtone can be purely five degrees pure octave relationship with the pitch, and the increase of these two overtones can improve the harmony of the arpeggio tone, reduce the hollowness of the sound, and enhance the intensity of the sound to make the sound more textured.

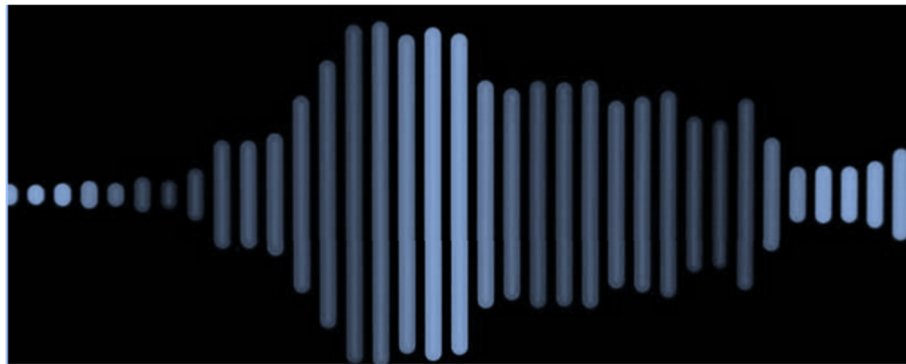
It can be seen from Figs. 7 and 8 that the pitch of the pitch is not in a state of sharpness, indicating that the singing process uses a more reasonable vibrato technique. At the same time, we can see the obvious second, third, and even fourth overtones after the pitch. Rich and obvious overtones increase the texture and harmony of the sound, while enhancing the brightness of the sound. Both the pitch and the overtone do not appear spiked, but appear to have a certain width "base," indicating that the singing state has a more reasonable vibrato. When the frequency is around 3200 Hz, the singer's formant has a strong peak, and its amplitude energy value is even at the fundamental frequency. This shows that the singing voice has a strong penetrating power and is not easily masked by the accompaniment of the symphony orchestra. The singer's mid- and low-range sounds are thick and sturdy. At the same time, the singer's first overtone, second overtone, and third overtone sound energy are significantly improved, adding more harmonious components to the sound timbre [16].

As can be seen from Figs. 9 and 10, in contrast to the bright tone of the soprano, the tone of the mezzo-soprano is relatively rich and heavy, which can be clearly reflected from the spectrum image analysis.



**Fig. 9** The a sound image of small group that Horne uses [o] vowel singing





**Fig. 10** The b sound image of small group that Horne uses [i] vowel singing

Similar to the soprano, the pitch and overtone of the mezzo-soprano are not the peak state, but the “base” has a certain width, which proves that the application of vibrato during singing is reasonable. The amplitude is between a small second, and the frequency of change is about six times per second. Similarly, the singer formant near 3200 Hz is evident in the spectral envelope, and the formant intensity energy is at the same level as the pitch sound energy. In contrast to the perception of the auditory system, the mezzo-soprano sound is thick, and although it does not have a bright sound quality, it still has a very strong penetrating energy, which fully reduces the acoustic masking effect of the band accompaniment. Similar to the soprano, the first overtone, the second overtone, the third overtone, and the fourth overtone of the mezzo-soprano are strongly magnified. However, unlike the soprano, the mezzo-soprano vocal cord is longer, which is more conducive to the third overtone and the fourth overtone. As I mentioned before, from the perspective of music and acoustics, the chord formed by the pitch and the first four overtones is just a major chord. While increasing the degree of harmony, it strengthens the distinction between the soprano tone.

## 5 Conclusion

Influencing factors for studying vocal vocalization, this study explores the new ideas and observation angles of traditional singing and vocal research through the analysis of sound spectrum data. Through spectrum image analysis, the singing voice can be digitized and quantified, and the abstract sounds that are difficult to ponder are presented in front of our eyes in the form of data charts. At the same time, through the good singing voice of typical singers as a relative reference standard, the vocal learners’ deficiencies in a certain aspect can be clearly compared. These shortcomings can be refined to each parameter, such as the pitch, tone, time value, and intensity of the vocal learner’s concern, and also clearly reflect the pitch, intensity, etc., of each overtone. In

addition, it can get more clear feedback on the envelope of the singer formant. For the spectrum image analysis of singing, its advantages are accuracy and visibility. However, it also has certain disadvantages. For singing art, over-emphasizing dataization will inevitably lead to the loss of artistic personality. This requires that the single spectrum data should not be used as the main basis for the evaluation mechanism, and it should be under the premise of “people-oriented.” At the same time, spectral image analysis is required as an auxiliary reference. Otherwise, it will inevitably have a bad influence. Research shows that this research method has certain practicality and can provide theoretical reference for related research.

### Acknowledgements

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

### Funding

Not applicable.

### Availability of data and materials

Please contact author for data requests.

### Author’s contributions

The author takes part in the discussion of the work described in this paper. The author read and approved the final manuscript.

### Competing interests

The authors’ declare that he have no competing interests.

## Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 October 2018 Accepted: 18 December 2018

Published online: 07 January 2019

### References

1. Xu G, Fowlkes J B, Tao C, et al. Photoacoustic spectrum analysis for microstructure characterization in biological tissue: analytical model[J]. *Appl. Phys. Lett.*, 2015, 41(5):1473–1480
2. Wang X, Xu G, Carson P. Quantification of tissue texture with photoacoustic spectrum analysis[J]. *Proc. SPIE - Int. Soc. Optical Eng.*, 2014, 9129:91291L

3. Hassani H, Heravi S, Zhigljavsky A. Forecasting UK industrial production with multivariate singular spectrum analysis[J]. *J. Forecast.*, 2013, 32(5):395–408
4. Guan H, Xiao B, Zhou J, et al. Fast dimension reduction for document classification based on imprecise spectrum analysis[J]. *Inf. Sci.*, 2013, 222(3): 147–162
5. Xu G, Meng Z X, Lin J D, et al. The functional pitch of an organ: quantification of tissue texture with photoacoustic spectrum analysis[J]. *Radiology*, 2014, 271(1):248
6. X. Qiu, P. Zhang, J. Wei, et al., Defect classification by pulsed eddy current technique in con-casting slabs based on spectrum analysis and wavelet decomposition[J]. *Sensors. Actuators. Phys.* **203**(12), 272–281 (2013)
7. Claudio M R S. Singular spectrum analysis and forecasting of failure time series[J]. *Reliab. Eng. Syst. Saf.*, 2013, 114(6):126–136
8. Zabalza J, Ren J, Wang Z, et al. Singular spectrum analysis for effective feature extraction in hyperspectral imaging[J]. *IEEE. Geosci. Remote. Sensing. Lett.*, 2014, 11(11):1886–1890
9. Fu K, Qu J, Chai Y, et al. Hilbert marginal spectrum analysis for automatic seizure detection in EEG signals[J]. *Biomed. Signal. Proc. Control.*, 2015, 18: 179–185
10. Chen Y. A wave-spectrum analysis of urban population density: entropy, fractal, and spatial localization[J]. *Disc. Dynamics. Nat. Soc.*, 2014, 2008(4):47–58
11. CHooper G. Nicholas Cook, *Beyond the score: music as performance* (Oxford: Oxford University Press, 2013). xiv + 458 pp. £32.99 (hb). ISBN 978-0-19-935740-6.[J]. *Music. Anal.*, 2016, 35(3):407–416
12. Kora S, Lim B B L, Wolf J. A hands-free music score turner using Google glass[J]. *J. Comput. Inf. Syst.*, 2017(5):1–11
13. Fang Y, Teng G F. Visual music score detection with unsupervised feature learning method based on K-means[J]. *Int. J. Mach. Learn. Cybernet.*, 2015, 6(2):277–287
14. Byo J L. Applying score analysis to a rehearsal pedagogy of expressive performance[J]. *Music. Educ. J.*, 2014, 101(2):76–82
15. Fine P A, Wise K J, Goldemberg R, et al. Performing musicians' understanding of the terms "mental practice" and "score analysis"[J]. *Psychomusicology*, 2015, 25:69–82
16. Louboutin C, Meredith D. Using general-purpose compression algorithms for music analysis[J]. *J. New. Music. Res.*, 2016, 45(1):1–16

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)