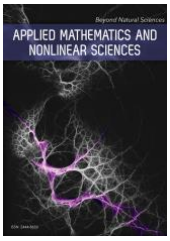
sciendo Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-19



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

The Application of Spectrogram in the Teaching of High-level Vocal Music Major Students

Juncheng Fang[1,†]

1. Conservatory of Music, Sichuan University of Science & Engineering, Zigong, Sichuan, 643000, China.

Submission Info

Communicated by Z. Sabir
Received April 27, 2024
Accepted August 9, 2024
Available online September 3, 2024

Abstract

Vocal music teaching has developed to the present day, still following the method of oral transmission, but with the development of modern science and technology, sound images can be processed to clearly see the quality of sound used. Students' vocal signals are used in the article to extract articulation features and construct a vocal spectrogram. The vocal spectrogram is used in vocal music teaching to enhance the students' timbre resonance. In order to verify its application in dissecting students' singing effects, the state of different students' American soprano articulation area is concretized on the spectrogram in terms of students' voice change differences, breathing aspects, and vocal resonance while comparing the differences between the manual evaluation and the evaluation of the vocal spectrogram analysis. It can be seen that the minimum fundamental frequency of student A in the first group is 191.7 lower than that of student B, which is 287.8, and the maximum fundamental frequency of 1033.2 is also significantly higher than that of student B, which is just the same as the conclusion drawn from the manual scale. To sum up, according to the sound spectrogram, the teacher can well analyze the students' voice waveforms visually, and the students can adjust the vocal state through the voice waveforms so as to carry out the correct reinforcement training and add new means for the traditional mode of vocal teaching, so as to make the teaching of vocal music gradually realize the visualization and intuition.

Keywords: Vocal music teaching; Sound spectrogram; Resonance peak; Articulation characteristics.
AMS 2010 codes: 68M11

†Corresponding author.
Email address: F13880184178@126.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2024-2475

1 Introduction

Vocal music teaching requires students to have not only singing skills but also the ability to interpret the work. The purpose of students' singing is to express the content of the song through their voices, while their voices express the emotion of singing. Therefore, students should not only have solid vocal skills and certain knowledge of vocal theory when singing but also have a strong ability to interpret and be able to interpret vocal works with rich and diverse emotions [1-2]. In vocal music teaching, teachers should not only teach students vocal skills and theoretical knowledge but also pay attention to cultivating students to show emotion in the singing process.

In the early stage of vocal music teaching, students are generally unable to understand the structure of the human organogenesis, there are doubts about how to scientifically vocalize, and the understanding of their voice is generally known through the objective evaluation of others, resulting in the inability to know the good and bad of their tone, and can not effectively regulate and control [3]. However, with the continuous development of computers, computer technology is becoming more and more developed, and the computer processing speed has been greatly improved, which provides favorable conditions for the development of the application of vocal spectrum analysis technology in vocal music teaching [4].

Using advanced sound spectrum analysis technology, the usual abstract sound concepts are transformed into computer graphics so that students can understand their voices more vividly and concretely, breaking through the traditional "one-to-one" teaching mode and realizing the "mouth-ear-nose" hybrid teaching mode [5]. Sound visualization focuses on the use of graphics to map the different frequencies and amplitudes of collected sounds, representing the interaction between the frequency components and the overall acoustics. The use of advanced computer technology is able to add all kinds of fun to the teaching classroom using text, video, images, and sound to fully stimulate students' passion for learning. Therefore, advanced computerized acoustic spectral analysis technology has been widely used in the classroom of vocal music teaching and a good results [6].

Literature [7] conceptualized a vocal teaching index assessment framework with pattern recognition vocal signature analysis as the core logic, which has a certain degree of accuracy by extracting and identifying the vocal signature of students' vocal occurrences in teaching and can be used for the assessment and analysis of students' vocal performance. Literature [8] describes the practice of vocal spectral analysis technology in vocal music teaching at the present stage while exploring the performance of GMM-SVM and DBN at the present stage and proposes a new convolutional neural network to extract the deep fusion features of vocal vocalization, which demonstrates the good accuracy of vocal pattern recognition in simulation experiments, which is positively significant for the development of vocal music teaching. Literature [9] carried out teaching practice to analyze the effect of pharyngeal training in vocal music teaching, combined with audio extraction technology and sound spectral analysis methods, pointed out that pharyngeal training is different from the special characteristics of general vocal music teaching, i.e., it requires large-scale and large-capacity training. Literature [10] proposed an audio spectrum analysis model with BP neural network algorithm and fast Fourier transform algorithm as the basic structure and introduced the Nios II system and cyclone IV to improve the compatibility of the model, which reached more than 95% of the audio spectrum resolution of vocal music in numerical tests, and has great potential for application in the field of music. Literature [11] discussed the principle of deep learning and the basic architecture of neural networks in the field of music recognition, and through empirical analysis, it was confirmed that deep learning technology can effectively improve the accuracy of sound recognition. Literature [12] carries out a comparative analysis experiment based on spectral image analysis technology and FFT algorithm to discover the characteristics and mechanisms of vocal voicing under different conditions

in vocal music teaching, which is an important reference value for the practice and research of vocal music teaching and vocal music training.

Vocal spectrograms are used to conduct a sonic study of voice majors. The article begins the analysis with a real-life case study where 45 students majoring in music performance were selected for this test and analysis. Firstly, five students were selected to identify their voice change differences using resonance peaks and vocal spectrograms. Secondly, some of the students were selected to begin the vocalization of the a-vowel in the high register of the American voice in terms of both respiration and vocal resonance. In the aspect of breathing, the vocal spectrogram was used to analyze the articulation of both thoracic and thoracic cases, and in the aspect of vocal resonance, the three different articulation cases were used to visualize the three different resonance peaks cases according to the vocal spectrogram. Finally, 10 students were selected to sing songs, comparing the manual evaluation with the evaluation based on sonograms to highlight the advantages of using sonograms in the vocal music program.

2 Recognition of student vocal events based on vocal spectrograms

2.1 Feature Extraction of Students' Vocal Pronunciation Characteristics

2.1.1 Student vocal signals in frames and windows

It is generally recognized that audio signals, including speech [13], are often highly time-varying and can be considered smooth only within a short period. Therefore, before performing feature extraction of acoustic parameters on audio signals, it is necessary to perform a frame-splitting operation, i.e., the signal is intercepted one segment at a time so that the signal is approximately smooth within each segment. The speech signal is generally 20-30ms in a frame. In order to remove the boundary effect, adjacent frames should overlap each other, and the commonly used overlap ratio is 1/2. Framing is generally realized by multiplying the original audio signal with a window function, and the commonly used window functions are rectangular window and Hamming window, which are mathematically described as (where N is the frame length):

Rectangular window:

$$w(n)=\begin{cases}1, & 0\le n\le (N-1)\\ 0, & n=\text{Other value}\end{cases} \tag{1}$$

Hamming windows:

$$w(n)=\begin{cases}0.54-0.46\cos[2\pi/(N-1)], & 0\le n\le (N-1)\\ 0, & n=\text{Other value}\end{cases} \tag{2}$$

2.1.2 Student vocal signal feature extraction

Feature extraction [14] is a key issue in pattern recognition. In general, good features should have the following properties:

1) Good inter-class differentiation, that is, these features can well reflect the differences between different classes.

2) Small intra-class distance, that is, the difference of the feature within a class should be as small as possible.

3) Good noise resistance, i.e., these features are not sensitive to noise.

4) The dimensions of the features should not be too high. Too high a dimension will increase the complexity of training the classifier and may overfit the model.

5) The correlation between features is as low as possible. Many researchers have devoted themselves to the work on feature extraction, and the currently available features can be roughly categorized into the following classes:

   (1) Time-domain features, such as short-time energy, short-time over-zero rate, and short-time average amplitude.

   (2) Frequency domain features, such as spectral center of mass, bandwidth, spectral roll-off coefficient, etc.

   (3) Cepstrum features, such as the Mel frequency cepstrum coefficient (MFCC), linear prediction cepstrum coefficient (LPCC), etc.

   (4) Time-frequency features, such as acoustic spectrogram features or wavelet transform features.

   (5) Underlay description features defined in the new MPEG-7 standard.

2.1.3 Dimensionality reduction of student vocal signal features

1) Calculate the eigenvalues and eigenvectors of Q

$$Qe_i = \lambda_i e_i \tag{3}$$

Where λ_i is the eigenvalue and e_i is its corresponding eigenvector, the eigenvalues are sorted from largest to smallest, and the eigenvectors corresponding to the first *m* eigenvalues as column vectors form the PCA linear transformation matrix of $n \times m$.

2) Dimensionality reduction of test data. In order to ensure consistency, the test data also need to be feature-centered before dimensionality reduction, i.e., each dimension of its data are subtracted from the mean value of the dimension of the training samples, and let the data at this time be y_i, then the new eigenvectors after dimensionality reduction can be expressed as:

$$z_i = W_{pca}^r y_i \tag{4}$$

n-dimensional before dimensionality reduction and *m*-dimensional after dimensionality reduction $(m << n)$. PCA can greatly reduce the feature dimensionality, but it is an unsupervised dimensionality reduction technique that does not take into account the category information and is not conducive to classification tasks.

3) Linear Discriminant Analysis (LDA)

Unlike PCA, LDA is a supervised dimensionality reduction technique that takes into account the category information between different kinds, and its goal is to make the inter-class scatter of the samples in the low-dimensional space after projection larger and the intra-class scatter smaller, which is more favorable for classification. Two evaluation criteria are defined, intra-class scatter matrix and inter-class scatter matrix, both of which are described mathematically as follows:

(1) Intra-class scatter matrix

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^r \tag{5}$$

where c is the number of categories, N_j is the number of samples of category j, x_i^j is the ith sample of category j, and μ_j, is the mean of category j.

(2) Inter-class scatter matrix

$$S_b = \sum_{j=1}^{c} N_j (\mu_j - \mu)(\mu_j - \mu)^T \tag{6}$$

μ is the average of all categories.

Let the transformation matrix of LDA be W. This matrix transforms the original m-dimensional column vector x into an f-dimensional column vector y, which is expressed by the following equation:

$$y = W^T x \tag{7}$$

The representation of a sample x obtained in the transformed space is y. It is easy to show that the intra- and inter-class scatter matrices of a sample in the transformed space are defined as follows:

$$S_w = W^T S_w W \tag{8}$$

$$S_b = W^T S_b W \tag{9}$$

If S_w is a non-singular matrix, to simultaneously maximize the interclass scatter matrix and minimize the intraclass scatter matrix, one can define the criterion function for the optimization as the ratio of the determinant of the interclass scatter matrix and the intraclass scatter matrix and maximize that ratio, i.e.:

$$J(w) = \frac{|S_b|}{|S_w|} = \frac{|\ W^T S_b W|}{|\ W^T S_w W|} \tag{10}$$

It is easy to show that the column vectors of the transformation matrix W that maximizes the criterion function consists of the eigenvectors corresponding to the largest eigenvalues in the following equation

$$S_b w_i = \lambda_i S_w w_i \tag{11}$$

This is a generalized eigenvalue problem, and if S_w is non-singular, then one can multiply both sides of the equation in Eq. (11) by S_w^{-1} to get

$$S_w^{-1}S_b w_i = \lambda_i w_i \tag{12}$$

The essence of equation (12) is to find the eigenvalues of the matrix. First, find the eigenvalues of matrix $S_w^{-1}S_b$, and then select the eigenvectors corresponding to the largest f eigenvalues as the column vectors to form the transformation matrix W. Next, we can use this matrix W to downscale the test data. It should be noted that the dimension of the subspace after LDA downscaling is at most $c-1$.

2.2 Acoustic Spectrogram Generation and Extraction

2.2.1 Generation of acoustic spectrograms

An acoustic spectrogram [15] is actually a spectrogram over time, a visual depiction formed by greyscale mapping and color mapping, which consists of three dimensions of information: frequency, time, and amplitude. The steps for generating the acoustic spectrogram are shown in Fig. 1. To transform a time-domain sound signal into an acoustic spectrogram, it is first necessary to do a short-time discrete Fourier transform of the signal:

$$X_i(k) = \sum_{n=0}^{N-1} x(n)w(n)e^{-\frac{2\pi i}{N}kn} \quad k = 0.....N-1 \tag{13}$$

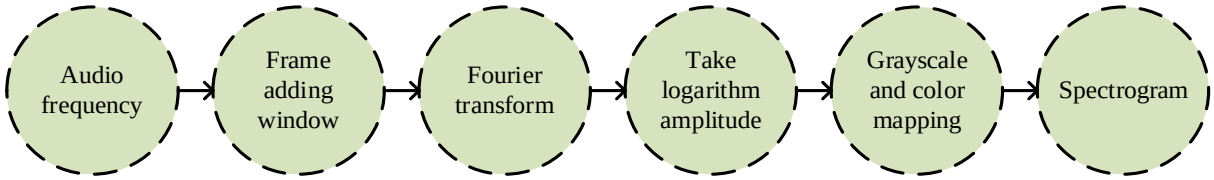


Figure 1. Process for generating sound spectra

where X represents the window length, $w(n)$ is the Hamming window function, k corresponds to a frequency of kf_i / N, and f_i is the sampling frequency. The acoustic spectrogram is then generated with a logarithmic amplitude description:

$$S(k,t) = 20 \times \log_{10} | X_i(k) | \tag{14}$$

2.2.2 Extraction of the main image

1) Independent Component Analysis (ICA)

ICA391 is a tool for processing and analyzing high-dimensional data using statistical principles and a classical sparse coding method. The method is used to separate data or signals into statistically independent non-Gaussian components of the source through multidirectional signal processing. The general procedure is to find a demixing system B based on observation X under the assumption that the components in the source S are independent of each other and both the source S and the mixing system A are unknown, such that the observation X: y output

after passing through this demixing system *B* is an optimal approximation to the source *S*. When considering the system as a matrix, the mixing matrix *A* is then the inverse matrix of the unmixing matrix *B*. The procedure of the ICA algorithm is shown in Figure 2 below:

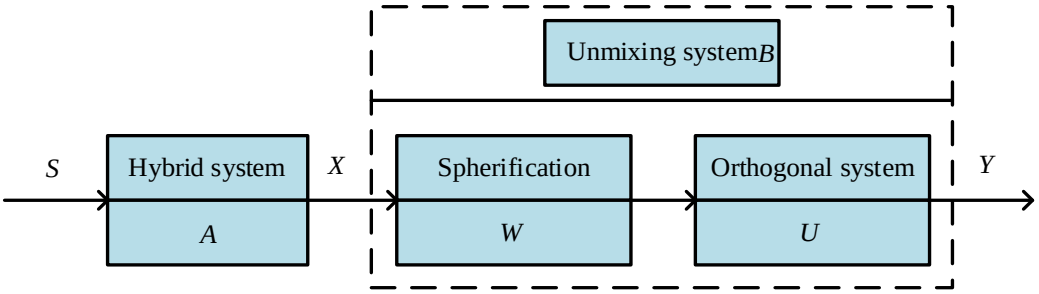


Figure 2. Process of independent component analysis algorithm

The process in Fig. 2 does not take noise into account, and the solution system B is composed of two parts, sphericalization and orthogonal system, where sphericalization, also known as whitening, is a linear system used to linearly transform the input data into the output data that is orthogonally normalized for each component.

2) Principal Component Analysis

The process of Principal Component Analysis (PCA) is to extract the main components of the data and rotate the data according to the primary and secondary components of the data, rearranged accordingly to remove the data after a few dimensions of the minor components of the data, high-dimensional indivisible data into low-dimensional data can be divided to achieve the dimensionality of the data. Specific PCA algorithm steps are as follows:

Suppose the i st image out of n images is x_i . First find the mean μ of the data and perform a zero mean operation of data normalization on the images by mean μ . The mean value and normalization are:

$$\mu = (\sum_{i=1}^{n} x_i) / n \tag{15}$$

$$x_i = x_i - \mu, i \in [1, n] \tag{16}$$

Then the covariance matrix A_{cov} of the image is computed and the primary and secondary directions of the image are derived by finding the eigenvector matrix V of the covariance matrix A_{cov} . Where the eigenvector matrix V needs to satisfy orthogonality, V_1 is the primary eigenvector, V_2 is the secondary eigenvector and so on in the matrix V. The covariance is calculated with its corresponding eigenvector matrix as:

$$A_{Cov} = (\sum_{i=1}^{n} x_i x_i^T) / n \tag{17}$$

$$V = [V_1, V_2, \cdots, V_n] \tag{18}$$

The obtained data is rotated and the original *n*-dimensional image data is projected into the new *n*-data x_r based on the primary and secondary directions of the feature vectors accordingly, and the projection formula is:

$$x_r = V^T x \tag{19}$$

The *m*-dimensional data after dimensionality reduction is obtained based on the primary and secondary directions of the new data x_r: $x_m = x_{r,m}$, where $x_{r,m}$ is the first *m*-dimensional matrix of x_r. When selecting the dimension *m* after dimensionality reduction, in order to avoid the data compression rate is not high due to *m* is too large or the data approximation error produced by *m* is too small, the empirical value of the variance of the image retention is used as a constraint, that is:

$$\sum_{i=1}^{m} \lambda_i / \sum_{i=1}^{n} \lambda_i \geq 0.99 \tag{20}$$

where λ is the eigenvalue of eigenvector V_i in equation (20) above.

Hierarchical contrast map which has the following two basic properties:

1) The energy of the sound source is most prominent, and the color feature is red.

2) Non-red areas are generally background noise or derivatives of the source energy. Therefore, a hierarchical contrast map is defined according to these two basic characteristics of the main map:

If the main map of the ird acoustic spectrogram is SM_i, the hierarchical contrast map SMR_i is defined as in equation (21):

$$SMR_i = \begin{cases} 0, if R(SM_i)/255 < 1 \\ 1, \text{other} \end{cases} \tag{21}$$

Equation (21) indicates that the hierarchical contrast map SMR_i is formed by normalizing and denoising the *R*-base color map $R(SM_i)$ of the main map SM_i, and the *R*-base color map $R(SM_i)$ is the red map in the three primary color color map of the main map SM_i.

2.2.3 Vocal Signal Recognition for Students Based on Vocal Spectrograms

A convolutional neural network is abbreviated as CNN. From the perspective of deep learning, CNN is the first learning algorithm that successfully trains a multilayer network structure, which belongs to the discriminative deep structure among the three deep structures of deep learning: generative, discriminative, and hybrid. In view of the advantages that convolutional neural networks can obviously perceive the local space with a strong correlation in natural images, this paper adopts LeNet-type convolutional neural networks to study the characteristics of sound sources in acoustic spectrograms. The algorithm can be studied at two major levels: the microscopic neuron and the macroscopic network structure:

Assume that a $m \times n$ large size image x is given, from which a $i \times j$ small size is extracted and set as the size of the mask y. $(i \le m, j \le n)$ Calculate the coefficient parameters w of the sparse filter in the network layer and set them to k different masks y. Convolve the image x with one of the masks y and add the bias b of the sparse filter, and finally obtain a characteristic map H of the response by a nonlinear function f with a displacement invariant function, such as a sigmoid or a tanh. The specific formula is as follows:

$$H = f(x * y + b) = f\left(\sum_{i=1}^{k} w_i x_i + b\right) \tag{22}$$

If k different features are extracted, there is a total of $k \times (m-i+1) \times (n-j+1)$ feature value and k feature maps after convolution, and the size of each feature map is $(m-i+1) \times (n-j+1)$.

Considering the specificity of the structure in the sound source image, for this reason, this section improves the LeNet-type convolutional neural network, whose structure is shown in Fig. 3.

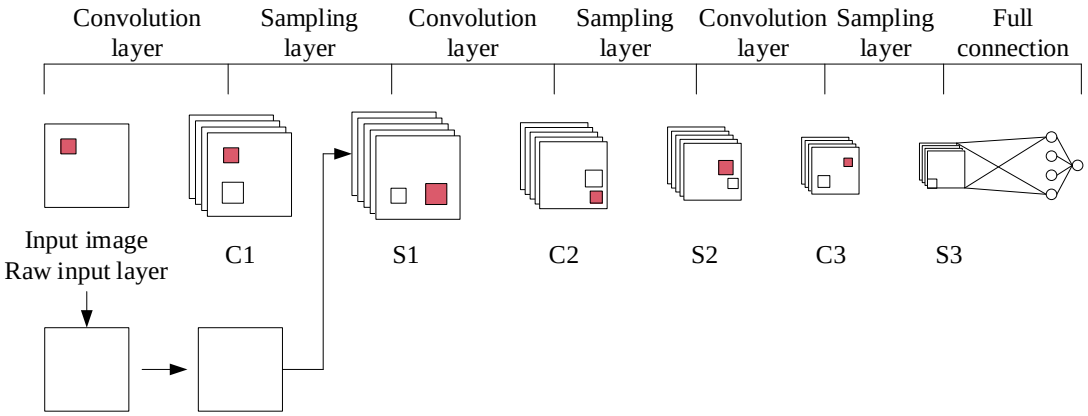


Figure 3. Improved convolutional neural network structure

The specific steps of the algorithm for audio image recognition using an improved convolutional neural network are as follows:

Assume that there is $m \times n$ known acoustic spectrogram $D = \{d_1, d_2, \cdots, d_i, \cdots d_{m \times n}\}$. Which encompasses n different sound sources with m acoustic spectrograms for each source;

Step 1: Extract the feature set using the feature extraction method: the set of hierarchical comparison maps $\mathrm{SMR} = \{SMR_1, SMR_2, \cdots, SMR_i, \cdots, SMR_{m \times n}\}$ with the set of feature maps $\mathrm{SMRP} = \{SMRP_1, SMRP_2, \cdots, SMRP_i, \cdots, SMRP_{m \times n}\}$ of PCA.

Step 2: Using the set of hierarchical contrast maps $\mathrm{SMR} = \{SMR_1, SMR_2, \cdots, SMR_i, \cdots, SMR_{m \times n}\}$ as the external input and the set of PCA feature maps $\mathrm{SMR} = \{SMR_1, SMR_2, \cdots, SMR_i, \cdots, SMR_{m \times n}\}$ as the internal input, training and modeling are performed by the improved deep convolutional neural network GCNN, so as to obtain the GCNN sound source models for n different sound sources;

Step 3: Identify the sound sources to be tested using the GCNN sound source models for n different sound sources Figure d';

Step 4: Obtain a hierarchical comparison map $SMR_{d'}$ and a PCA feature map $SMRP_{d'}$ of the acoustic spectrogram d' to be tested, following the method of extracting features in step 1.

Step 5: The hierarchical contrast map $SMR_{d'}$ of the sound spectrogram d' to be tested is used as the input image of the external input layer; the PCA feature map $SMRP_{d'}$ of the sound spectrogram d' to be tested is used as the input image of the internal input layer; so as to recognize the sound source to which the sound spectrogram d' to be tested belongs by using the GCNN sound source model of n different sound sources.

Since the hierarchical contrast map is a simplified master map, it can itself characterize the structure of the sound source and, thus the overall characteristics of the source. The overall characteristics of the sound source must be taken into account in the recognition of the hierarchical contrast map, and it is important to analyze the sound source. Therefore, the improved convolutional neural network is more suitable for source image and feature recognition of audio spectrograms and further improves the accuracy of the algorithm, even though it only adds a new internal input part to the internal principle structure.

3 Spectrogram technology-driven vocal instruction

Spectral analysis technology has been widely used in vocal music teaching, which not only fundamentally improves the overall effect of vocal music teaching but also plays an important significance in the smooth development of teaching activities. Therefore, it has been favored by many vocal music teachers.

3.1 Traditional vocal instruction

The sound spectrum is a graphic representation from which we can not only obtain the amplitude of different frequencies, but also have a clear understanding of the components of each frequency. With the rapid development of China's computer technology and network technology, its application in vocal music teaching is also more and more extensive, which also promotes the improvement of the sound spectrum analysis technology to a certain extent. The improvement of technology can be better reflected through the image of the quality of the voice, to sing in the breath as an example, excellent singers in the process of singing the sound spectrum, not only the depth of the obvious but also enough breath, even and coherent. Even and coherent. The resonance of the voice can last longer, or vice versa. Therefore, using the vocal spectrum to study the singing breath, we can use aerodynamics to complete the study of the depth and volume of breath, and at the same time, we can also have a comprehensive grasp of the factors that can have an impact on the voice in each state, so as to better control the quality of the voice. However. In the traditional teaching mode of vocal music, the judgment of the quality of the voice relies entirely on the sense of hearing and can not be fully integrated into the visual role, which will inevitably lead to a certain extent to the quality of the voice of the judgment of the existence of errors, and can not play a scientific reference role. However, the addition of spectral analysis technology can effectively combine vision and hearing to analyze the quality of the singer's voice and summarize the problems that occur in the singing process, which has a more obvious teaching effect and, thus, more conducive to the teacher's teaching and students' understanding of their voices, and can greatly improve the quality of teaching.

3.2 Application of Sound Spectrum Technology in Teaching and Learning

In the course of teaching. Teachers can use the sound spectrum analysis technology to effectively assess the noise of each student so as to take targeted intensive training for the actual situation of students so that the learning efficiency of students can be further improved. In addition, the use of acoustic spectral analysis technology is essential for student voice therapy. The pitch and fundamental frequency of the singer can be effectively grasped through the vocal spectrum parameters, and with an increase in pitch, the loudness of the voice also increases. The gray scale of each resonance peak shows the performance on the speech spectrum. Increasing the use of resonance versus not using resonance. There is a difference in the speech spectrum, as shown in the use of resonance. The loudness of the second and third resonance peaks significantly increases. Professional vocalizations using resonance compared to non-professional vocalizations using resonance were more pronounced at the second resonance peak. It can be seen that the application of vocal spectral analysis technology to vocal music teaching can not only better improve the students' deficiencies but also improve the quality of teaching to a large extent and promote the sustainable development of vocal music arts and culture.

4 Example analysis of the application of vocal charts to the teaching of vocal music

To verify the effectiveness of vocal charts in teaching and learning, this paper selects 45 majors in the music performance department of a school for vocal analysis based on vocal charts. Separately, students were examined for their recognition of voice change, vocal ability, and timbre.

4.1 Student vocal prosody recognition

4.1.1 Resonance peaks

The human body vocalizes from the vocal cords to the lips to escape from this sound channel has a number of important resonance cavities. When the sound source in the overtone frequency and the resonance cavity itself are close to the frequency of the sound, it will become stronger and louder. In a certain frequency range in the spectral envelope, there will be towering peaks and valleys called resonance peaks.

Student resonance peak is 2500 ~ 4000hz overtones. This band of overtones ensures that the sound penetration, in fact, is a professional student familiar with the high position of the resonance. This band of overtones is an excellent voice professional students must have. This frequency region of overtones ensures that the song penetrates the accompaniment.

Boy's resonance peak range is generally 2500Hz~3200Hz, while the girl's for 2800Hz~3500Hz. Although men's and women's vocal structures are different, and singing the male spectrum pitch and the actual pitch difference of eight degrees, the resonance peak range is relatively similar. Therefore, it is easy to think that the vocal cords and the air column only produce the fundamental overtones, while other factors, such as the shape of the vocal tract and the use of the cavity determine the resonance peaks of the voice.

4.1.2 Analysis of sound spectral differences between student voice changes

The purpose of this experiment with a few selected students from a group of 45 students was to verify that those frequency bands were present when they produced color shifts in their prosodic voices.

The e^2 open transposition sung by Student 1 with the vowel is shown in Figure 4, and this spectrogram highlights several of the characteristics:

1) At the same normal sound pressure level, the fundamental is much less loud than the other overtones, which are already below the -1db scale in the illustration.

2) It is the first six overtones that affect the timbre of the human voice, and these overtones are very prominent in this student's voice, indicating that he is skillful enough to utilize the resonance of the various chambers.

3) This student has a very typical singer's resonance peak in the frequency range of 2000 to 6800 Hz, which is a common characteristic of good vocal practitioners.

However, open transitions are most prominent in this band of resonance peaks, which are significantly higher in amplitude than all other frequencies, followed by the first overtone and the second and third overtones. The fundamental, which is usually higher than all other frequency bands in similar and precession experiments, is the lowest in the range of the "fundamental and first six overtones combination", which has the greatest effect on the color of the human voice. Therefore, this type of voice modification is less burdensome on the human vocal cords, and the singer is less prone to vocal fatigue, which saves more energy.

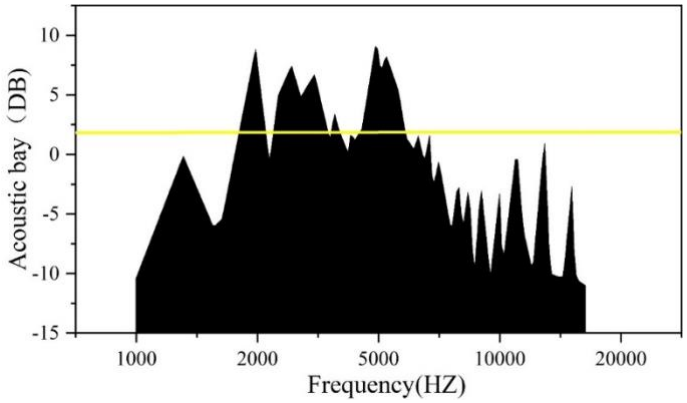


Figure 4. Open change sound

The differences between the different methods of voice change for two students of the same level at the same pitch and the same sound pressure level are shown in Figure 5. In the analysis, it was concluded that the pitch and intensity of the fundamental and first overtones remained unchanged, but the second overtones (2000 Hz, sound name b3) were significantly enhanced, the third overtones and fourth overtones were reduced, and the sixth overtones still maintained a high percentage. This reflects the characteristics that distinguish closed singing from open voice change:

1) The elevation of the first and second overtones and the lowering of the fifth and sixth overtones are commonly shown in this inverse proportion in many experiments.

2) The first and second overtones affect the brightness and darkness of the timbre and have the most obvious effect on the quality of the sound; the increase in the amplitude of the fifth and sixth overtones significantly increases the brightness of the timbre and at the same time affects the audible height of the sound position and a small portion of the spatial sense of the role.

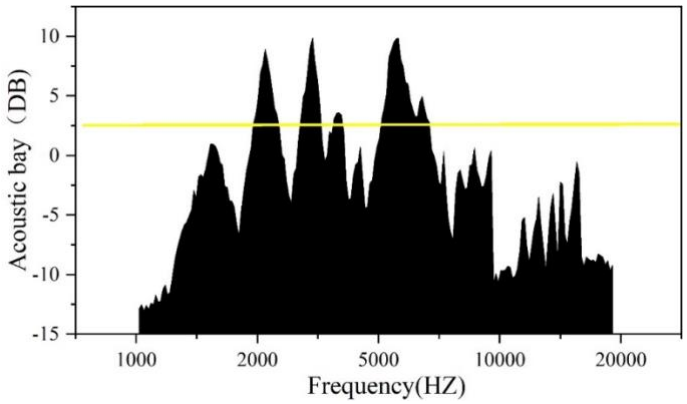


Figure 6. Closed singing

Student 4 sings e^2 openly with the vowel, as shown in Figure 6, with the base note similarly below the first, second, fifth, and sixth overtones.

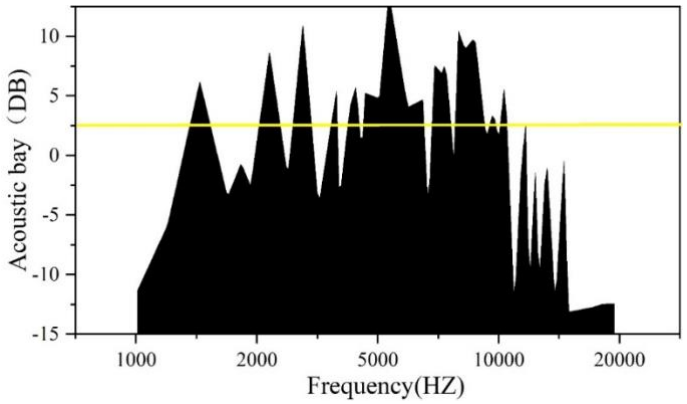


Figure 6. Open change sound

Through the observation of the overall loudness on the decline, as shown in Figure 7, off the sound change: in the sixth overtone above the overtone drop is more obvious, the base and other overtones roughly down 5db. But the obvious difference is that the singer resonance peaks in the frequency band of 2200 ~ 6600Hz almost did not reduce. This indicates that sound concentration should be maintained under both methods of voice change.

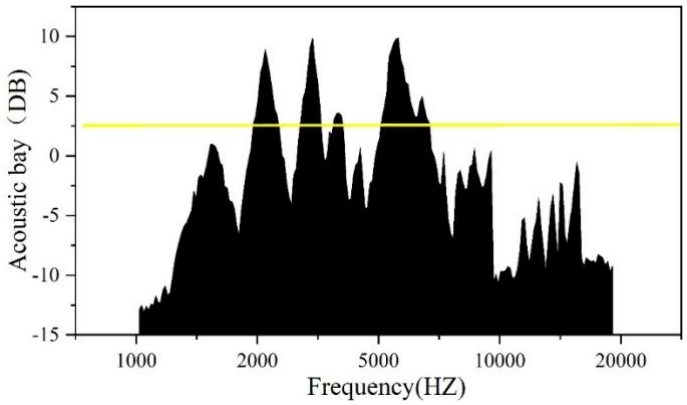


Figure 7. Turn off change sound

4.2 Analysis of students' vocal ability

This section specifically will be through the auditory system on the timbre of the perception mechanism, the use of spectral measurement means, the use of student resonance peaks, sound pressure, the number of overtones, and other parameters of the student's singing voice measurements and analysis of the sound samples into visible and measurable graphic signals, the transformation of the abstract sound for the intuitive graphic, so that the vocal teacher in the student's voice based on the auditory judgment.

4.2.1 Respiratory aspects of students

Breathing is the foundation of singing and can be categorized into chest breathing, abdominal breathing, and combined chest and abdominal breathing. In vocal music teaching, the more common method for beginners is thoracic breathing, also known as clavicular breathing. This kind of breathing singers relies on the lateral expansion of the rib cage to inhale, driving the clavicle elevated, breath inhalation shallow, less capacity, the diaphragm's movement is smaller, making the diaphragm and abdominal muscles lose the ability to control the breath, breath from the upper chest out, easy to make the neck of the laryngeal muscles tense, affecting the expansion of the range of the voice and the unity of the vocal range, resulting in a dry, lack of change in the voice. Thoracic and abdominal joint breathing is the thorax and diaphragm compression of air-filled lungs so that the breath is gradually out of the body. This method of using the thorax, diaphragm, and abdominal muscles to jointly control the breath is currently the vocal community that is a more scientific and reasonable method of using the breath for most teachers and students. It can fully mobilize the dynamic role of the human respiratory organs, enhance the ability to control and support the voice, the scope of respiratory activity, and the volume of the breath to increase the volume of the breath, scalability for the breath balanced, smooth exhaled to provide the conditions, but also stabilize the two ribs and diaphragm tension, to facilitate the grasp of the larynx downward discharge of the gas pressure changes, to avoid the impact of the breath is too strong to make the sound of the voice high, low, strong, weak changes freely, tone It makes the voice high, low, strong, weak and change freely, and the tone is round, loud, rigid and soft, with strong penetrating power.

In the following, the difference in spectrograms between chest breathing (Fig. 8) and chest-abdominal combined breathing vocalizations (Fig. 9) will be observed specifically by an American singing soprano in the same vocal range (high register) for the same vowel (a-vowel):

A comparison of the spectrograms of chest breathing vocalizations with those of the chest-abdominal breathing method shows that:

1) The spectrogram of vocalization with chest breathing shows 3 overtones, which is obviously less. There is no obvious song resonance peak in the frequency range of 3700-4300Hz. The height of the resonance peak is relatively low, indicating that the vocalization is dry and lacks change and penetrating power.

2) The number of overtones shown on the spectrogram of the vocalization with the chest and abdominal breathing method is 4. There are 2 obvious resonance peaks between 2100-2800Hz, with a certain peak width and peak height. There is another resonance peak in the frequency band after that, with the frequency range of 3100-3400Hz, which shows that the vocalization has more overtones and more Appears in the high-frequency range. The tone is round, bright, and penetrating.

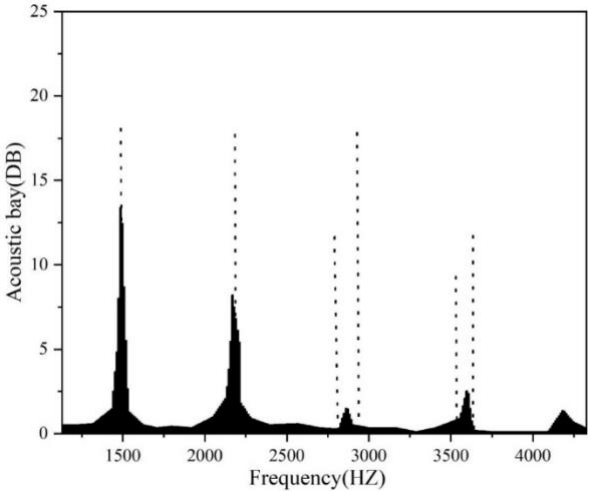


Figure 8. A graph of the sound spectrum of a vowel sound in a student's sound

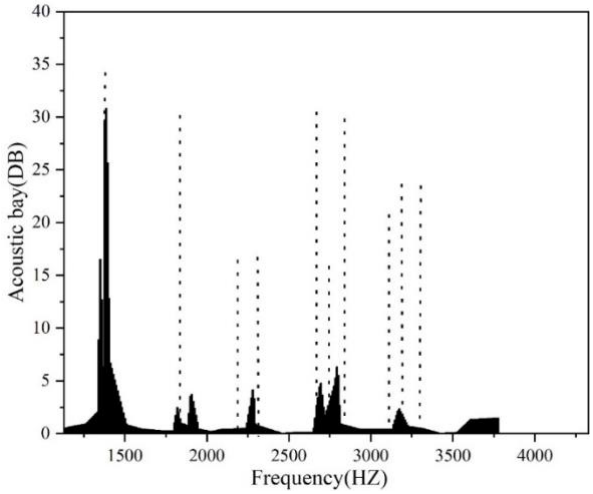


Figure 9. A student's sound voice spectrum

4.2.2 Aspects of students' vocal empathy

A more common problem with resonance in singers is low resonance position and multiple oral resonances. For convenience, we observed the difference in the spectrogram by comparing the correct resonant vocalization of the vowel in the high soprano register of a student (Fig. 10) with nasal resonance (Fig. 11) and out-of-place position (Fig. 12).

From the above comparison of the spectrograms of correct vocalizations with the spectrograms of nasal resonant, out-of-place vocalizations, it can be seen that:

1) The correct resonance vocalization of vowel a in the high register of a student's American voice is shown in Figure 10, which shows a higher number of overtones, 4, and 2 distinct resonance peaks in the frequency range of 2500-3500 Hz, with the resonance peaks in the frequency range of 2300-3000 Hz having a certain width and height and being more dense, and in the frequency range of The resonance peak in the frequency domain of 2300-3000Hz has a certain width and height, and is relatively dense, and there is also 1 obvious resonance peak in the frequency domain of 3400-3800Hz after 3000Hz, which indicates that the sound is relatively mellow, bright and rich in penetrating power.

2) The frequency spectrum of a student's a-vowel nasal sound in the high register of the American voice is shown in Figure 11. The figure shows a moderate number of overtones and three and two-singer resonance peaks in the frequency range of 2500-3400Hz. Still, the peak height change is not obvious enough. The peak range is not wide enough. The density of resonance peaks is relatively sparse compared to the correct vocalization, indicating that the voice has a certain degree of brightness, but the penetrating power is slightly insufficient.

3) The frequency spectrum of a student's vocalization in the high register of the American voice with the vowel out of place is shown in Figure 12. The figure shows that the number of overtones is obviously low: 1. In the frequency range of 2500-3000Hz, there are no obvious resonance peaks of the singer, indicating that the vocalization obviously lacks penetration, relatively dry.

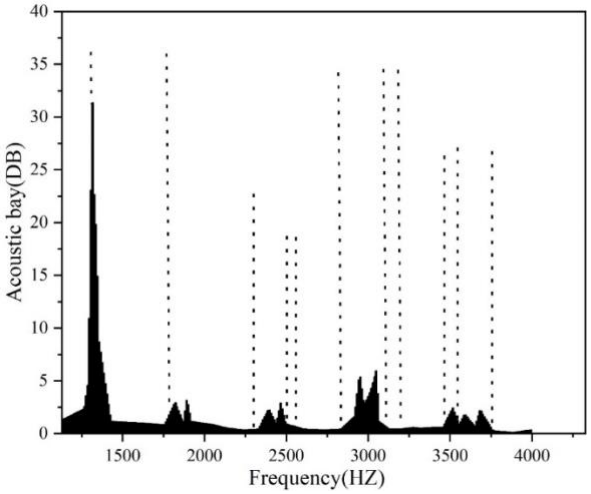


Figure 10. A sound spectrum of a vowel sound in a student's sound

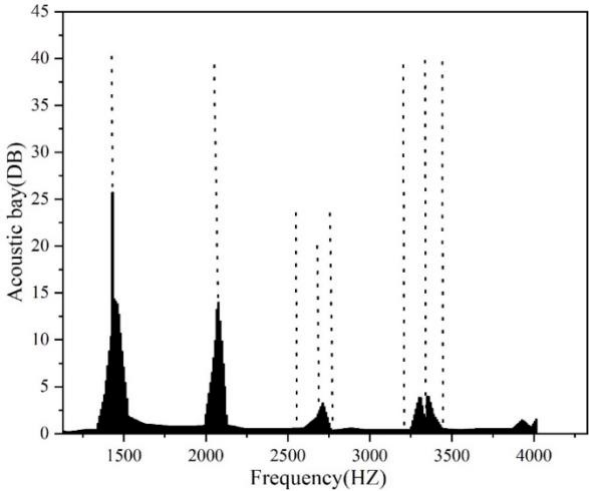


Figure 11. A student's sound voice spectrum map of a vowel sound

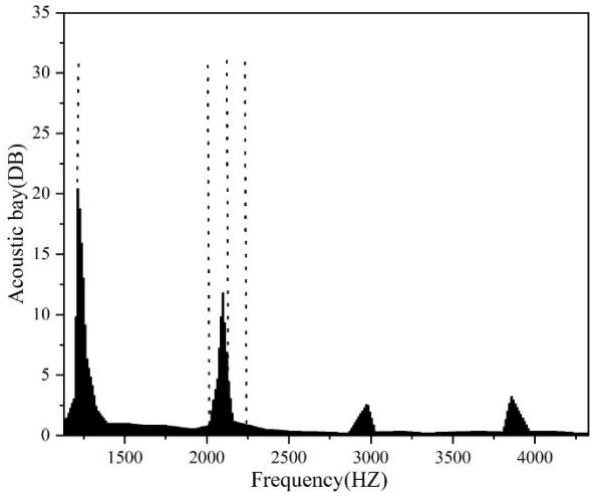


Figure 12. A student's sound noise area a vowel is not a sound spectrum

4.3 A study of student timbre based on sound spectral techniques

In this section, from the 45 students who performed music, 10 girls who performed better were selected to sing "Little Bridge and Flowing Water" as an example, and the students were first evaluated in terms of timbre using manual evaluation and then evaluated using the sound-spectrogram. The manual evaluation is shown in Table 1, and the comparison results of the sound spectrogram are shown in Table 2. In Table 2, F4 represents the frequency of the word "bright", followed by F0-avg, which represents the average fundamental frequency of the audio; F0-min, which represents the minimum fundamental frequency; F0-max, which represents the maximum fundamental frequency, STN, which is the number of semitones, and HNR, which is the number of semitones, and HNR, which is the number of semitones, and STN, which is the number of semitones. Number of semitones, and HNR is the harmonic-to-noise ratio in dB.

According to Table 1, we can conclude that most of the 10 second-year female students majoring in vocal music have a good foundation for vocal music study. Still, it can also be seen that there are some problems with individual students' vocal singing techniques, for example, the number of

choices describing the voice as wide, loose, and solid, the breath as deep, the timbre as bright, forward and rounded in the data displayed is relatively small. The number of evaluations of the timbre as backward is 44, and the coherence has a relatively small number. The number was 44, coherent was 31, and shallow breath was 40. This indicates that the overall sound of the students is still backward, not coherent enough, and their breath is still shallow. However, some female students can achieve a more stable tone and breath, but these are only the results of subjective data from the audience, after which the author will combine with the spectral analysis to obtain further experimental data and then analyze it.

Table 1. The audience evaluates the statistical tables

Semantics	Track "Small bridge"	
Breadth and narrow	24	41
Loose and tight	16	31
Real and virtual	39	44
Deep and shallow	30	40
Bright and dull	30	39
Lean back	13	44
Roundness and acuity	23	52
Coherence and disconnect	31	27
Flow and stagnation	27	38
Flexible and dull	31	34

The minimum fundamental frequency of student A is lower than that of student B, and the maximum fundamental frequency is significantly higher than that of student B, as shown in Table 2. This is just the same as the conclusion drawn from the manual scale, that A's voice is more smooth, while B's voice has the problem of uneven highs and lows. The maximum fundamental frequency of A is higher than that of B by nearly 300 Hz, and the minimum fundamental frequency of A is significantly lower than that of B by 300 Hz. In the manual evaluation, listeners generally think that A has a wider range and is more relaxed and mellow. B300Hz, during the manual evaluation, the listener generally agreed that the sound range of A is wider, with a more relaxed, rounded, and bright sound. After further observation, it is found that in any group of comparators, the value of HNR, the smaller the negative value, the higher the degree of artificial evaluation of its timbre evaluation, and vice versa, the larger the value of HNR, the artificial evaluation is relatively low. For example, in the first group of results, the Harmonic Noise Ratio (HNR) index of No. A is lower than that of No. B students and the comparison result of HNR in the second group is -9.6:-6.5; the comparison result of HNR in the third group is -5.4:-4.7; and the comparison result of HNR in the fourth group is -3.6:-3.4; and -3.4:-2.2 for the fifth group of HNR comparisons. From the listener's manual evaluations and spectral analyses, the larger the F0-max value, the brighter the subject's timbre, and if there is a broken tone, the value of STN increases accordingly.

Based on the above, it was concluded that the smaller the minimum fundamental frequency usually behaves, the larger the exponent of the maximum fundamental frequency has to be, and the lower the harmonic-to-noise ratio exponent of the subjects, their voices generally tended to be what we would recognize as the better voices, and the sound spectra responded well to the student's timbre.

Table 2. The sound spectrum results compare the table

Group	Student	F4	Fo-avg	Fo-min	Fo-max	STN	HNR
First set	A	4088	311.6	191.7	1033.2	29	-9.4dB
	B	3125	503.4	287.8	733.2	16	-4.3dB
Second group	C	4021	475.1	123.1	992.5	36	-9.6dB
	D	3811	522.4	278.2	751.2	23	-6.5dB
Third group p	E	4096	483.5	204.2	1006.6	31	-5.4dB
	F	3544	526.1	232	762.3	24	-4.7dB
Fourth group	J	3533	480.4	65.7	988	50	-3.6dB
	H	3012	522.6	203.5	927.6	28	-3.4dB
fifth grou	J	3357	476.1	206.7	992.3	28	-3.4dB
	K	3275	505.5	156.3	756.3	24	-2.2dB

5 Conclusion

In this paper, we analyze the differences in transcription between students' timbres using vocal spectra. We then compare these transcriptions in practice by displaying breathing, resonance, and other vocalizations of students' a-vowel articulations in the high register of the American voice on a spectrogram. Finally, we analyze the students' singing using both manual and spectral evaluations. In summary, we arrived at the following conclusions:

1) The spectrogram can well reflect the singing status of different students, and it is obvious that the students who sing with open vocalization are the most excellent among them by singing with voice change. At the same time, closing the singing will affect the listener's sense of hearing and spatial perception.

2) Through the spectrogram and resonance peaks, the shortcomings and strengths of different students' vowel articulations are shown. The vocal spectrogram of the student who pronounced the vowels correctly in the high register of the American voice responded that the number of overtones of the student was 4, 2 resonance peaks appeared in the frequency range of 2500-3500Hz, and 1 obvious resonance peak appeared after 3000Hz, and its frequency range was between 3400-3800Hz.

3) Students sang "Little Bridge and Flowing Water" based on manual evaluation and sound spectrogram evaluation, and the results showed that the sound spectrogram technique was consistent with the manual evaluation results.

The sound spectrogram can transform sound into a shape, allowing students to not only hear their voices during vocal practice but also visually observe their voice shape and adjust their vocal state through visual analysis of the sound waveform for targeted training.

References

[1] Li, J. (2020). Feasibility Study on the Application of the Computer Visualized Audio Parameter Analysis Method in the Vocal Music Digital Teaching. In Frontier Computing: Theory, Technologies and Applications (FC 2019) 8 (pp. 1520-1525). Springer Singapore.

[2] Jiang, K. (2023). An Introduction to the Innovative Path and Exploration of the Teaching Reform of Ethnic Vocal Music. Applied Mathematics and Nonlinear Sciences, 9(1).

[3] Wang, B. (2023). Visual transmission algorithm of vocal music resources based on sound spectrum analysis technology. Reviews of Adhesion and Adhesives, 11(2).

[4] McQuade, M. (2020). Dynamic Uses of Spectrographic Analysis in Choral Rehearsals and the Voice Studio. Journal of the Association for Technology in Music Instruction, 1(1), 1.

[5] Chi, X. (2017). Study on vocal music teaching innovation mode based on computer simulation and voice spectrogram analysis. Revista de la Facultad de Ingenieria, 32(16), 400-406.

[6] Cui, X., & Chen, M. (2024). A novel learning framework for vocal music education: an exploration of convolutional neural networks and pluralistic learning approaches. Soft Computing, 28(4), 3533-3553.

[7] Chen, N. (2021, September). Vocal music teaching evaluation model based on pattern recognition and voiceprint feature analysis. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 2800-2804).

[8] Zhang, X. (2021). Research on Modeling of Vocal State Duration Based on Spectrogram Analysis. In E3S Web of Conferences (Vol. 236, p. 04043). EDP Sciences.

[9] Huang, C. (2022). Vocal music teaching pharyngeal training method based on audio extraction by big data analysis. Wireless Communications and Mobile Computing, 2022(1), 4572904.

[10] Hao, J. (2022). Optimizing the design of a vocal teaching platform based on big data feature analysis of the audio spectrum. Wireless Communications and Mobile Computing, 2022(1), 9972223.

[11] Liu, N. (2022). Study on the Application of Improved Audio Recognition Technology Based on Deep Learning in Vocal Music Teaching. Mathematical Problems in Engineering, 2022(1), 1002105.

[12] Sun, J. (2019). Research on vocal sounding based on spectrum image analysis. EURASIP Journal on Image and Video Processing, 2019, 1-10.

[13] Alexandru Madalin Vizitiu,Lidia Dobrescu,Bogdan Catalin Trip,Vlad Florian Butnariu,Cristian Molder & Simona Viorica Halunga.(2024).Detection of the Compromising Audio Signal by Analyzing Its AM Demodulated Spectrum.Symmetry(2),

[14] Ma Mengzhen,Hu Ying,He Liang & Huang Hao.(2024).GLFER-Net: a polyphonic sound source localization and detection network based on global-local feature extraction and recalibration.EURASIP Journal on Audio, Speech, and Music Processing(1),

[15] AbayomiAlli Olusola O.,Damaševičius Robertas,Abbasi Aaqif Afzaal & Maskeliūnas Rytis.(2022). Detection of COVID-19 from Deep Breathing Sounds Using Sound Spectrum with Image Augmentation and Deep Learning Techniques.Electronics(16),2520-2520.