International Journal for Housing Science and Its Applications

Publish September 9, 2025. Volume 47, Issue 1 Pages 432–442

https://doi.org/10.70517/ijhsa47136

Analysis of Vocal Cord Vibration Characteristics in Modern Popular Vocal Performance Techniques

Zhiyong Li[1,*]

[1] School of Music, Xinxiang University, Xinxiang, Henan, 453003, China

Corresponding authors: (e-mail: lizhiyong2988318@126.com).

Abstract This paper employs traditional acoustic features such as fundamental frequency, frequency perturbation, amplitude perturbation, Mel-frequency cepstral coefficients (MFCCs), and linear predictive cepstral coefficients (LPCCs) to study the identification of vocal characteristics in modern pop singing. By combining dynamic vocal fold image sequences, the vocal fold vibration components are extracted. Through the selection of key points on the vocal folds and the extraction of feature parameters, the dynamic analysis of vocal fold vibration characteristics under different singing techniques is conducted. There are significant differences in vocal fold spectra under different singing techniques. For example, in the growling singing state, there are more sound impurities and poorer tone purity, while in the normal chest voice technique, overtones are strong and account for a larger proportion. The glottal area change index under the normal chest voice technique ranges from 0.379 to 0.437, significantly lower than that of the growling and breathy voice techniques, but higher than that of the uniformly distributed vibration state. Additionally, the average vibration rate at three key vocal fold points is 3500–4500, with an amplitude of 5–11, showing significant differences from growling, localized vibration, and breathy voice techniques. The vocal fold vibration characteristics of modern pop vocal techniques exhibit high diversity and individuality, pointing to new research directions in music science.

Index Terms acoustic feature extraction, vocal fold image sequence, vibration feature parameters, modern pop vocal music, singing technique

I. Introduction

The singing techniques of popular vocal music have evolved with the times and have been influenced by various factors throughout their development [1], [2]. Overall, the development of popular vocal music has been a winding path, but its future is bright. Its singing techniques are increasingly diversifying and presenting a distinct image to audiences. Vocal fold vibration characteristics are a key factor influencing the performance of modern popular vocal music [3]-[6].

Popular singing techniques emphasize natural, flexible, and efficient breath support [7]. Diaphragmatic breathing is key, as the contraction and expansion of the diaphragm enable more thorough inhalation and control of breath. Good breath control ensures sufficient inhalation volume to support sustained vocalization for 8–15 seconds or longer [8]–[11]. In fast-paced songs, the inhalation frequency increases appropriately, but each inhalation must still provide sufficient support for short, compact notes [12]-[14]. In pop singing, the vocal placement is more forward, concentrated in the front of the mouth, particularly the anterior region of the hard palate [15], [16]. By focusing the sound in this area, the timbre becomes brighter, clearer, and more penetrating [17]. During vocalization, the vocal fold vibration frequency range is broad. In the low-pitched range, the vocal fold vibration frequency typically ranges from 80 to 250 Hz, producing a deep, rich timbre [18], [19]. In the mid-range, the frequency ranges from 250 to 1000 Hz, resulting in a smooth, natural timbre [20]. In the high-pitched range, frequencies can reach 1000–4000 Hz or higher, producing a bright, sharp, and highly expressive sound [21].

Popular singing techniques encourage singers to showcase their unique personal style and rich emotional expression, with variations in sound intensity, clarity, and brightness that align with the emotional tone of the song [22], [23]. When expressing tender, heartfelt emotions, the voice can be appropriately softened, with volume controlled between 40–60 decibels, to convey emotions through a gentle, delicate tone [24], [25]. When expressing passionate, uplifting emotions, volume can be increased to 80–100 decibels, making the voice more powerful and explosive [26], [27]. Additionally, by controlling breath and vocalization techniques, one can achieve variations in vocal texture, such as using a 30–50% ratio of “soft voice” in lyrical sections to create a dreamy, emotional atmosphere [28], [29]. In climactic sections, the ratio of “hard voice” can be increased to 70–90% to enhance the power and impact of the voice [30].

This study effectively extracted traditional acoustic features such as fundamental frequency, frequency perturbation,

amplitude perturbation, Mel-frequency cepstral coefficients, and linear predictive cepstral coefficients during modern pop vocal performance. Additionally, based on dynamic vocal fold image sequences, it obtained key points and feature parameters of vocal fold vibration. Additionally, several learners and performers of modern pop music were invited to establish a vocal fold signal database for different singing techniques according to the established experimental procedures. Based on this, effective analysis was conducted on the vocal fold movement state, vibration rate, and amplitude of the researchers.

II. Extraction and research design of vocal cord characteristics in modern popular vocal music performance

II. A. Traditional acoustic feature extraction

II. A. 1) Fundamental frequency, frequency perturbation, and amplitude perturbation

Fundamental frequency (FO): Refers to the fundamental frequency of vocal fold vibration, a basic parameter in the study of pathological voice, widely used in acoustic voice analysis. In this study, the autocorrelation function method was employed to extract the fundamental frequency at integer points of the fundamental frequency period. The autocorrelation function of the voice signal exhibits significant peaks at these points. Identifying the first largest peak allows estimation of the fundamental frequency period. Additionally, the fundamental frequency is the reciprocal of the fundamental frequency period.

Frequency Perturbation (FP): Refers to the small frequency variations between adjacent periods in human voice. Frequency perturbation can serve as an indicator for evaluating pathological voice. The parameter is primarily expressed in terms of the absolute value, relative value, and frequency perturbation quotient (PPQ) of frequency perturbation. The calculation formula is as follows:

$$Jita = \frac{1}{N-1}\sum_{i=1}^{N-1} | T_0^{'} - T_0^{i+1} | \tag{1}$$

$$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} | T_0^{(i)} - T_0^{'+1} |}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \tag{2}$$

$$PPQ = \frac{\frac{1}{N-4}\sum_{i=1}^{N-4} \left| \frac{1}{5}\sum_{r=0}^{4} T_0^{\{i+r\}} - T_0^{i-2} \right|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{\{i\}}} \tag{3}$$

In the above three equations, $T_0^i, i = 1,2,3,\cdots,N$ are the fundamental frequency periods extracted. N is the number of periods.

Amplitude perturbation refers to small changes in amplitude between adjacent periods of human voice. Its main forms include the absolute value, relative value, and ratio of amplitude perturbation. The calculation formula is as follows:

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1} 20\log(A^{(i-1)} / A^{i}) \tag{4}$$

$$Shim = \frac{\frac{1}{N}\sum_{i=1}^{N-1} | A^{(i)} - A^{(i-1)} |}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \tag{5}$$

$$APQ = \frac{\frac{1}{N-4}\sum_{i=1}^{N-4} \left| \frac{1}{5}\sum_{i=0}^{4} A^{(i+1)} - A^{(i-2)} \right|}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \tag{6}$$

In the above three equations, $A^{(i)}, i = 1,2,3,\cdots,N$ are the extracted peak amplitudes. N is the number of cycles.

II. A. 2) Mel frequency domain cepstrum coefficients

Mel-frequency cepstral coefficients (MFCC) are primarily based on the auditory characteristics of the human ear and are analyzed on the Mel scale [31], [32]. Since the frequency of sounds received by the human ear is nonlinearly proportional to

the frequency of sounds emitted by humans, the Mel frequency scale is more consistent with the auditory characteristics of the human ear. The relationship between Mel frequency and actual frequency is as follows:

$$Mel(f) = 2595\lg(1 + f / 700) \tag{7}$$

Among them, the unit of the actual frequency f is Hz.

II. A. 3) Linear predictive cepstrum coefficients

Linear predictive cepstral coefficients (LPCC) are features established based on vocal tract modeling, mainly reflecting the response characteristics of the vocal tract [33], [34]. The effect of the vocal tract on the vocal fold excitation signal can be regarded as a time-varying filter with all poles, whose system function is as follows:

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{8}$$

Among them, $a_1, a_2, \cdots, a_p$ are time-varying parameters.

Then, the difference equation of the voice signal is:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{9}$$

$s(n)$ is a linear combination of the first p voice signal sample values, where $a_1, a_2, \cdots, a_p$ are the linear prediction coefficients.

II. B. Acquisition of key point feature parameters based on dynamic vocal fold image sequences

II. B. 1) Extraction of vocal fold vibration components

In the raw vocal cord images collected through experiments, in addition to the vocal cords, there are also a significant number of other laryngeal muscles and organs. Edge detection refers to the process of extracting the contours of an image during digital image processing. To clearly understand the image and reduce interference from the local environment of the vocal cords on the user, the study employed the Sobel edge detection operator to perform edge detection on the vocal cord region, thereby obtaining the contour pattern of the vocal cords.

Gray-scale images are two-dimensional images, and their gradients can be divided into horizontal and vertical components. The Sobel operator combines Gaussian smoothing and differential calculus, incorporating two gradient matrices: the horizontal gradient matrix Gx and the vertical gradient matrix Gy, both of which are 3x3 in size [35]. The Sobel operator is an improved first-order differential operator, where the gradient matrices have been weighted in specific directions. Using this operator, the edge gradients of an image can be calculated, resulting in an image that displays the brightness changes along the edges. The formula is as follows:

$$Gx = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & -2 \\ -1 & 0 & 1 \end{bmatrix} * A, Gy = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \tag{10}$$

Among them, A is a digital image for edge detection, and the numerical value of the image gradient can be expressed as:

$$G = \sqrt{Gx^2 + Gy^2} \tag{11}$$

II. B. 2) Selection of key points on the vocal cords

Since the contour of the vocal cords is elliptical, the dynamic parameters of different points on the same side of the vocal cords will inevitably vary during vibration. To obtain vocal cord dynamic characteristic parameters that can be compared and analyzed, corresponding key points are selected on both sides of the vocal cords. When the vocal cords are closed, these corresponding key points come into contact with each other, meaning that the key points on both sides of the vocal cords are symmetrical. By extracting and comparing the dynamic parameters of the corresponding key points, the vibrational state of the vocal cords in each sample can be analyzed.

Figure 1 shows the key points selected on the vocal cords. According to relevant laryngological researchers, when the vocal cords are in operation, differences in vibration between the two sides of the vocal cords are more likely to occur at the 1/3 point, 2/3 point, and 1/2 point of the vocal cords. Therefore, these three points were selected as the key points of the vocal cords. By comparing the dynamic characteristic parameters on both sides of the key points, the state of the vocal cords of the

singer can be analyzed.

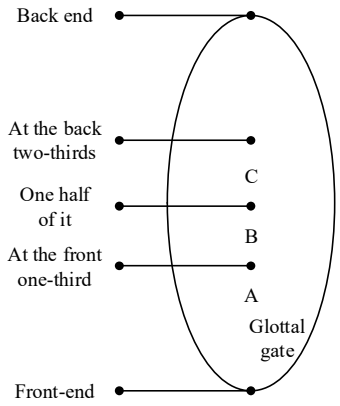

Figure 1: Key points of vocal cord selection

II. B. 3) Extraction of key point feature parameters of the vocal cords

During the dynamic process of vocal fold vibration, the primary characteristic parameters are the average velocity and amplitude of vocal fold vibration. In the raw image sequences collected during the experiment, in addition to the vocal folds, there are also many other laryngeal organs and muscles that cause interference. Additionally, factors such as liquid reflections around the target area and the angle of photography can also affect the extraction of vocal fold features.

To obtain an accurate set of vocal fold contour points and extract the dynamic features of vocal fold vibration, while simplifying the image processing workflow and minimizing the over-extraction of edges during edge detection, it is first necessary to obtain a Region of Interest (ROI) containing the vocal fold region as the effective image sequence for subsequent processing.

After obtaining the required ROI region, to improve the accuracy of target region information extraction and reduce the volume of raw data in subsequent image processing, the target image sequence can be converted to grayscale, thereby achieving the goal of reducing subsequent computational processing load.

To obtain an effective vocal fold vibration contour point set, thresholding processing must also be applied to generate a binary image containing the vocal folds, thereby eliminating the influence of the surrounding environment. The expression is as follows:

$$B(x,\ y,\ t)=\begin{cases}255, & (F(x,\ y,\ t)>\theta)\\ 0, & (\text{other})\end{cases} \tag{12}$$

Here, θ is taken as the binarization threshold.

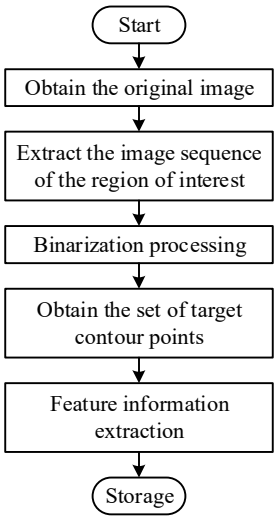

Figure 2: Flowchart of Feature Parameter extraction

After removing other interfering factors in the environment, a contour tree containing the contour point set of the vocal cord image can be obtained through image feature extraction, and the non-outer layer point set in the contour tree can be removed. The remaining point set is the required effective contour point set of the vocal cord vibration. Subsequently, the dynamic feature parameters of the corresponding points in the contour point set are obtained using the selected key points. The overall process of feature parameter extraction for dynamic vocal cord images is shown in Figure 2.

II. B. 4) Obtaining the average velocity and amplitude at key points of vocal fold vibration

When the human body produces sound, the vibration contour of the vocal cords is elliptical. Select the upper end point (x1, y1) and lower end point (x2, y2) of the complete vocal cords. Set the current key point coordinates to $c_n(X_n, Y_n)$ for each operation. As shown in the following formula. The x-axis and y-axis coordinates of each key point are as follows.

x-axis coordinate of key point A:

$$X_n = \frac{x1 - x2}{3} + x1 \tag{13}$$

Y-axis coordinate of key point A:

$$Y_n = \frac{y1 - y2}{3} + y2 \tag{14}$$

X-coordinate of key point B:

$$X_n = \frac{x1 - x2}{2} + x1 \tag{15}$$

Y-axis coordinate of key point B:

$$Y_n = \frac{y1 - y2}{2} + y2 \tag{16}$$

X-coordinate of key point A:

$$X_n = \frac{2*(x1 - x2)}{3} + x1 \tag{17}$$

Y-axis coordinate of key point A:

$$Y_n = \frac{2*(x1 - x2)}{3} + y2 \tag{18}$$

where n denotes the nth key point from the upper endpoint to the lower endpoint.

As can be seen from the formula, the key points are basically distributed along the central axis of the vocal cords. The two corresponding points on either side of the vocal cords will necessarily coincide at the key points when the vocal cords are closed. In the same sample, these two corresponding points can be used to describe the dynamic characteristic information of the vibrations on both sides of the vocal cords at the corresponding key points.

To obtain the dynamic image feature information of the corresponding points on both sides of the vocal folds at the key points, it is first necessary to obtain the coordinates of each corresponding point. Substitute the coordinates $N(x, y)$ of each point in the vocal fold contour into the following formula:

$$a = ((x1 - X_n), (y1 - Y_n)) \tag{19}$$

$$b = ((x - X_n), (y - Y_n)) \tag{20}$$

$$cos\alpha = \frac{a \cdot b}{|a| \cdot |b|} \tag{21}$$

where a is the vector from the upper endpoint A to the n th keypoint, b is the vector from the N points on the contour point set to the n th keypoint, and cos is the cosine of the angle between vectors a and b. By comparing the absolute values of $cos\alpha$, we can obtain the critical points C_n corresponding to the left and right sides of the vocal cords, respectively, $cos\alpha$ minimum points $L_i(x_i, y_i), R_i(x_i, Y_i)$ on the left and right sides of the vocal cords, respectively.

From the known corresponding points, the left amplitude A_L and right amplitude A_R corresponding to the key points can be calculated using the following formulas:

$$A_L = \sqrt{(x_i - X_n)^2 + (y_i - Y_n)^2} \tag{22}$$

$$A_R = \sqrt{(X_i - X_n)^2 + (Y_i - Y_n)^2} \tag{23}$$

By accumulating the displacement of the corresponding points throughout the entire vibration process, the average vibration rate of the corresponding points can be obtained. The formulas for the average vibration rate of the left vocal cord V_L and the average vibration rate of the right vocal cord V_R are as follows:

$$V_L = \frac{\sum_{i=1}^{M}(|c_n D_i| - |c_n D_{i+1}|)}{t \cdot M} \tag{24}$$

$$V_R = \frac{\sum_{i=1}^{M}(|c_n E_i| - |c_n E_{i+1}|)}{\tau \cdot M} \tag{25}$$

where M is the total number of frames in the entire motion sequence of the video, and t is the time interval between adjacent frames.

II. C. Research Design

Building on the research achievements of previous scholars, and by studying relevant knowledge of phonetic theory and drawing on the experience of other academic papers, we have designed an experimental study on the vocal fold vibration characteristics of modern popular vocal performance techniques.

II. C. 1) Experimental procedure

Ten professional modern pop vocalists, including both faculty members and students, were selected from M Music Academy in a certain city as the vocalists for this study. Vocal cord signal data were collected using different techniques in the academy's recording studio. The vocal techniques involved included: normal chest voice, growl, partial vibration, and breathy voice. Each technique was combined to form different vocal cord study conditions in this experiment. Among these, normal chest voice and growl techniques formed the vocal fold active force closure state (T1), while normal chest voice combined with full vibration and partial vibration techniques formed the vocal fold passive force closure state (T2). Normal chest voice combined with breathy voice techniques formed the vocal fold uneven force and non-closure state (T3).

Subsequently, in the preliminary preprocessing stage, the selection, recording, and segmentation of the speech data were completed, the file naming and saving rules were clarified, and a database was established. Next, parameters such as energy, formants, duration, fundamental frequency, root mean square, and velocity mean square were extracted and a database was established.

II. C. 2) Recording Equipment and Methods

The primary hardware equipment used for signal acquisition in this experiment includes a laptop computer, Audition recording software, a laryngoscope, an external sound card, a mixing console, and a microphone. The hardware system required for studying glottal impedance signals is the glottal impedance meter (EGG). During the use of the electronic glottal impedance meter, a pair of electronic sensors are fixed on either side of the thyroid cartilage, ensuring close contact with the thyroid cartilage. When the vocal cords vibrate, the flow rate undergoes amplitude adjustments. The collected signal (EGG signal) reflects changes in glottal impedance. Among these, three parameters are most important: fundamental frequency, opening quotient, and velocity quotient. Typically, while collecting glottal impedance signals, it is also necessary to collect speech signals; therefore, the hardware setup also requires a microphone, external sound card, and mixing console.

Glottal impedance signals are collected via EGG, while speech signals are collected via a microphone. The two-channel signals (glottal impedance signals and speech signals) are then input into the computer simultaneously through an external sound card via the mixing console. The final collected signal is a two-channel signal.

Save the recorded speech files, categorize them by topic, and store them in the same folder to form a corpus. Then use Cool Edit software to segment the recordings. Divide the entire song into sections, sentences, and words to prepare for subsequent analysis and research.

II. C. 3) Data extraction and statistical analysis

The parameters of speech signals are primarily extracted and analyzed using the Praat speech analysis software. The energy, duration, and formant parameters of long tones are extracted, and a comparative analysis is conducted on the energy, duration,

formant parameters, and other characteristics of long tones across three different genres to identify differences. The parameters of vocal signals include fundamental frequency, overtone, and velocity quotient, which can be used to describe and define different types of vocal production. In the study of speech production types, these definitions can be applied to different singing techniques and vocal styles in vocal music research. When extracting vocal signal parameters, the Real-Time EGG Analysis software on the Multi-Speech 3700 processing platform was utilized.

Opening quotient extraction steps: Open Real-Time EGG Analysis → load EGG data → options → select real-time EGG options → EGG quotient → open quotient → apply → OK → Numerical results → save as → file naming.

Speed quotient extraction steps: Open Real-Time EGG Analysis → load EGG data → options → select real-time EGG options → user defined → set user quotient → select opening phase and closing phase → OK → apply → confirm → numerical results → save as → file naming. This paper utilized Excel and SPSS 19.0 for data statistics, analysis, and graphing.

III. Analysis of vocal cord vibration characteristics in modern popular vocal performance techniques

III. A. Comparison of vocal fold movement spectra

The movement patterns of the vocal cords undergo certain changes under different force conditions. The primary points of observation include the degree of vocal cord closure, vibration state, and the movement patterns of the vocal cord muscles. This section investigates the vocal fold movement states under different techniques used in modern pop vocal performance, specifically including three states: active force-induced closure of the vocal folds (T1), passive force-induced closure of the vocal folds (T2), and uneven force application with incomplete closure of the vocal folds (T3). Using the methods described above, vocal fold characteristics were extracted under different states, and spectral analysis was conducted to preliminarily explore the vibrational movement characteristics of the vocal folds in modern popular vocal singing techniques.

When the human voice is affected by factors such as different vocal states, pitch, and intensity, it can interfere with the experimental results for the listener. Therefore, the study used the "General Music Analysis System GMAS (2.0B)" software to record and analyze the spectrum under the standard tone A3.

Active force-induced vocal fold closure state (T1):

Figure 3 shows the spectral comparison results between normal chest voice and growl techniques. Figure (a) is the spectrum recorded under the normal chest voice technique, where the vocal folds are in a relatively relaxed closed state. Figure (b) shows the vocalization state during the performance of a tear sound, where the vocal cords are in an active closed state. The peaks and valleys of the waveform represent the intensity of the overtone series. Comparing the two diagrams, it can be observed that when the voice is in an active closed state, the peaks and valleys of the waveform are more prominent, with noticeable and disordered fluctuations, and the waveform movement contains more "spikes," indicating a more complex sound quality. The waveform in the diagram fluctuates frequently, with an abundance of dissonant overtones, indicating a significant difference in timbre. This corresponds to the singer incorporating false vocal fold closure during the growl sound, resulting in a noticeable contrast in timbre compared to the true voice technique. In the true voice technique shown in the figure above, overtones are more pronounced and account for a larger proportion. The waveform peaks and valleys are orderly and not chaotic, indicating that the sound quality has fewer noises and the timbre purity is more prominent.

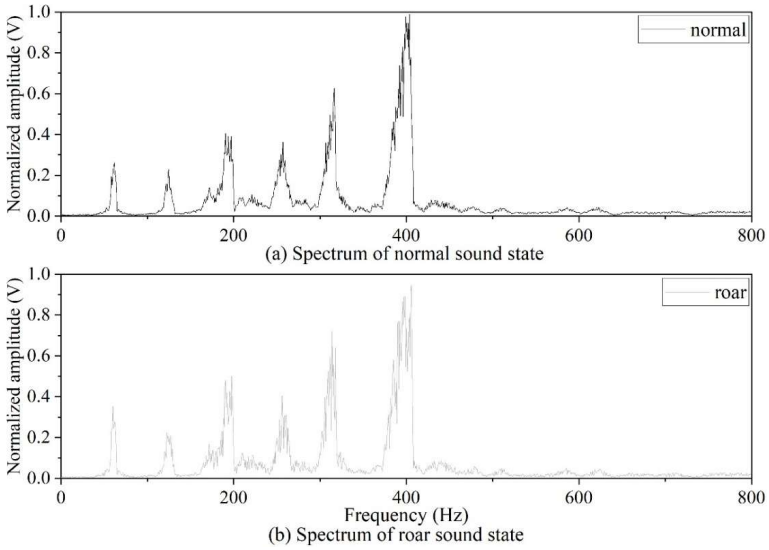

Figure 3: Normal sound is compared to the state spectrum of the roar

Vocal fold closure state under passive force (T2):

Figure 4 shows the spectral comparison results between complete vibration and partial vibration states of the vocal folds. As can be seen from the spectral comparison, when the vocal folds are in a partial vibration state, the waveform is simpler, with more distinct peaks and valleys, and a smoother overtone series, indicating a more relaxed vocal state. There are fewer dissonant harmonics, and the waveform maintenance is higher than that of normal true voice techniques, making the voice more recognizable. Additionally, compared to the spectrum of normal true voice techniques, there are fewer "spikes" in the waveform during partial vibration, resulting in higher voice stability and a softer tone.

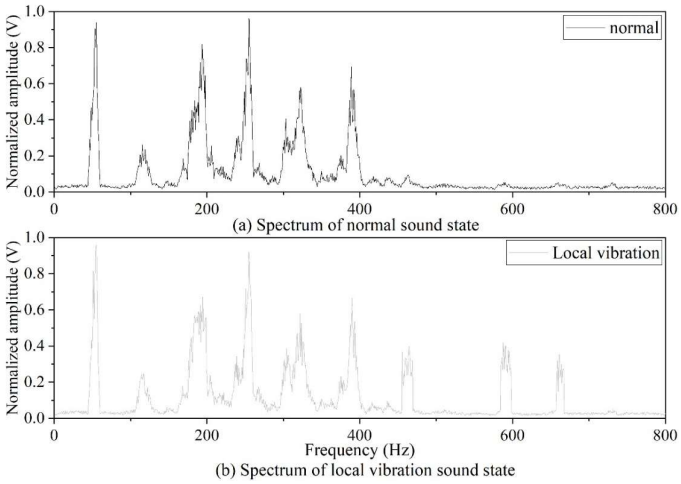

Figure 4: The vibration state spectrum of the vocal cords is compared

Uneven vocal fold loading and incomplete closure (T3)

The spectral comparison results between normal chest voice technique and breathy voice technique are shown in Figure 5. Comparing the spectral diagrams under the two different techniques, it can be observed that the waveform peaks and troughs in the spectrum are significantly reduced under the breathy voice technique. Except for the waveform peaks and troughs of the first fundamental frequency, which are nearly identical to those of the chest voice technique, the subsequent overtone series exhibit a cliff-like decline. During true voice singing, the intervals between overtones are small, with similar values, and each interval is followed by a similar overtone. In contrast, under the breathy voice technique, the overtone series only shows a significant response at the natural frequency band. This reflects that during breathy voice singing, the natural voice color is prominent and the tonal quality is distinct in the first half, but as the test progresses to the latter half, the breathy sound overpowers the natural voice volume, affecting the purity of the tone color, while the overall volume also decreases. In breathy singing, the area of vocal fold closure and vibration is smaller compared to true voice singing, resulting in a significant difference in timbre perception, with breathy singing sounding darker. However, unlike breathy singing, in breathy singing, the vocal cords are more relaxed, and the control ability of the vocal cord muscles further decreases, causing uneven force on both sides of the vocal cords as airflow passes through, resulting in incomplete vocal cord closure. In this case, the fundamental tone is largely masked by the strong airflow. As the degree of glottal closure increases, the vocal cord closure effect strengthens, and the tonal quality becomes more prominent. In breathy singing with good muscle control, the fundamental frequency is not masked by the breath and is accurately and clearly conveyed to the listener's subjective perception.

III. B. Glottal area change index

Analyzing high-speed laryngeal imaging of 10 researchers during singing to obtain the glottal area change index (Maxi-Mini)/Maxi under different singing techniques.

Figure 6 shows the glottal area variation index during pop vocal performance by the researchers. As shown in the figure: under normal chest voice technique, the glottal area variation index fluctuates between 0.379 and 0.437, with an average of 0.407 ± 0.015. Under the growling technique, the Maxi-Mini index fluctuates between 0.485 and 0.628, with an average of 0.558 ± 0.021, and the fluctuation range is significantly higher than that of the normal chest voice technique. The overall curve for normal individuals is smoother and more stable, with minimal fluctuations in glottal area during vocal fold vibration. In the local vibration state, the overall glottal area variation index fluctuates between 0.198 and 0.251 over 10 vocal fold vibration cycles, with an average of 0.229 ± 0.015, indicating a significantly smaller fluctuation range compared to the normal true voice technique. Under the breathy voice technique, the glottal area change index over 10 vocal fold vibration cycles ranged from 0.486 to 0.585, with an average of 0.526 ± 0.241, significantly higher than that of normal true voice technique. Overall, the glottal area change index showed no significant fluctuations across various singing states. This is because researchers adopted

a relatively stable vocalization method during singing, resulting in consistent glottal opening and closing amplitude.

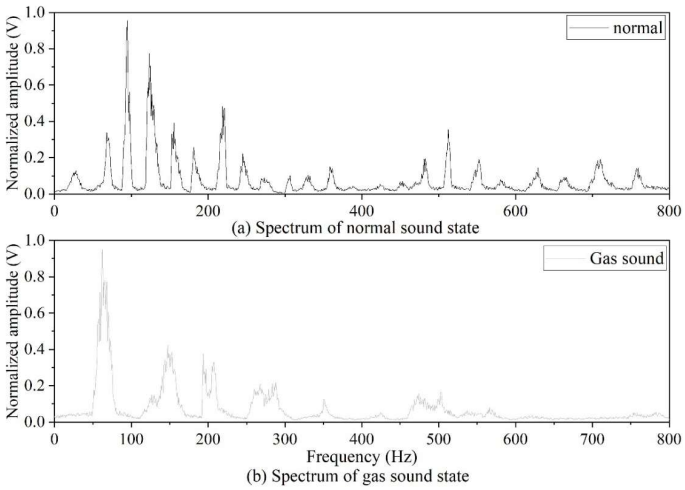

Figure 5: The spectral contrast between true sound state and sound state

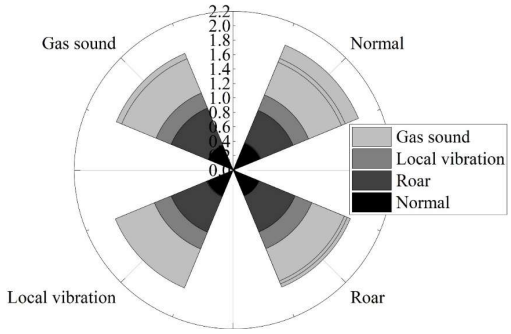

Figure 6: The sound gate area change index in popular vocal music

Additionally, the average speed quotient (SQ) values for each group were obtained through analysis of the glottal area change index, and the differences in SQ values between normal voice, shouting, local vibration, and breathy voice techniques were compared. A higher SQ value indicates a longer duration of the open phase during the glottal cycle, while a lower SQ value indicates a longer duration of the closed phase during the glottal cycle.

Table 1 lists the average SQ values for different singing states. As shown in the table, there are statistically significant differences between normal true voice and the comparison states in three conditions: active force closure (T1), passive force closure of the vocal cords (T2), and uneven force and non-closure of the vocal cords (T3). For example, the SQ value for normal chest voice in the active force closure state is 1.26 ± 0.28, which is significantly smaller than the SQ value for the shout state. This indicates that in the shout state, the glottal closure speed is significantly faster than the opening speed, leading to short, intense closure impacts. In the passive force closure state (T2), the SQ values of normal true voice and local vibration are similar, indicating that the duration of vocal fold opening movement during the glottal cycle is relatively similar in both states. However, in the uneven force and incomplete closure state (T3), the duration of opening movement during the glottal cycle in normal true voice is significantly longer than that in the breathy voice technique.

Table 1: Study the sq average of different singing states

	T1		T2		T3	
	Normal	Roar	Normal	Local vibration	Normal	Gas sound
SQ value	1.26±0.28	2.54±0.44	1.13±0.25	1.37±0.18	1.31±0.19	0.86±0.11
P	0.000		0.0431		0.000	

III. C. Average vibration rate and amplitude measurement

To further analyze the vibrational characteristics of the vocal cords under different conditions, this section utilized a high-

speed visual laryngoscope to obtain high-resolution, high-precision raw vibrational images of the vocal cords. Researchers then statistically analyzed the average vibration velocity at three key points (A, B, C) of the vocal cords under three distinct singing states: active force-induced closure (T1), vocal fold closure under passive force (T2), and vocal fold uneven force and non-closure (T3) under different singing states.

Table 2 presents the statistical results of average vibration rates and amplitudes of the vocal cords under different singing states. As shown in the table, under normal chest voice technique, the average rates of the three key points A, B, and C of the vocal cords, regardless of which side, range between 3,500 and 4,500, which exhibits significant differences compared to growling voice state, local vibration state, and breathy voice technique. This is because different singing techniques restrict the movement of the anterior, middle, and posterior key points of the vocal cords, thereby significantly affecting the vibration rates at these locations. In terms of amplitude comparison, changes in the average vibration rate of the vocal cords influence the average amplitude of the three key points (A, B, and C). As shown in the table, the average amplitude under normal chest voice conditions ranges from 7 to 10, indicating a moderate state. The average amplitude in the growl state ranges from 5 to 11, with both higher and lower amplitude values. In both the local vibration and breathy voice techniques, the average amplitude ranges from 6 to 9, indicating a relatively smaller average amplitude. This is because differences in glottal pressure, vocal fold gap, and muscle contraction intensity across various singing techniques result in significant differences in average amplitude at all three key points of the vocal folds, regardless of the side.

Table 2: The average vibration rate and amplitude statistics of the vocal cords

Project		T1		T2		T3	
		Normal	Roar	Normal	Local vibration	Normal	Gas sound
Left velocity	A	3869	2349	4335	4783	4428	1892
	B	3595	2494	4273	5166	4020	1640
	C	3771	2757	3495	4484	4430	1492
Right velocity	A	4126	2434	4044	5278	3708	2172
	B	4068	2494	4220	4944	4483	1817
	C	3994	2573	4010	5453	3455	2043
Left amplitude	A	8.713	9.378	10.623	7.402	8.616	8.836
	B	8.453	9.429	9.848	6.296	8.469	7.329
	C	9.314	5.298	10.781	8.417	9.963	8.709
Right amplitude	A	8.974	10.757	9.849	7.657	8.956	8.524
	B	7.745	9.052	8.862	8.597	9.047	8.065
	C	9.201	7.924	10.886	8.737	9.536	6.543

IV. Conclusion

This paper first applies traditional acoustic feature extraction methods to extract multiple acoustic features of vocal cord signals during modern pop vocal performances, including fundamental frequency and frequency fluctuations. Then, through steps such as extracting the vibrating portion of the vocal cords, selecting key points of the vocal cords, and extracting key point feature parameters, a dynamic vocal cord image sequence is obtained. Finally, 10 relevant personnel were invited to conduct experiments on vocal fold movement states, assisting in the analysis of vocal fold vibration characteristics in modern pop vocal performance techniques. When using the tear sound technique, the spectral fluctuations are larger, the sound quality is more complex, and the contrast with normal true voice timbre is significant. When the vocal folds are in a localized state, the waveform peaks are prominent, the sound is stable, and the timbre is gentle. Under the breathy voice technique, the number of waveform peaks and valleys in the spectrum is fewer than in normal true voice, and the subsequent overtone series shows a decreasing trend. The index of glottal area changes under different singing techniques exhibits fluctuating differences. The average index under the growl technique and breathy voice technique ranges from 0.485 to 0.628, significantly higher than that of normal true voice. Additionally, there are significant differences in the average movement rate and amplitude at three key points on the vocal cords between the two sides. The average rate under normal true voice technique ranges from 3,500 to 4,500, and the amplitude ranges from 7 to 10. Different singing techniques exhibit unique physiological characteristics in vocal cord vibration. Analyzing these vibration characteristics is of great significance for vocal pedagogy, voice health, and music production.

References

[1] Heidemann, K. (2016). A System for Describing Vocal Timbre in Popular Song. Music Theory Online, 22(1).

[2] Larrouy-Maestri, P., & Morsomme, D. (2014). Criteria and tools for objectively analysing the vocal accuracy of a popular song. Logopedics Phoniatrics Vocology, 39(1), 11-18.

[3] Liu, J., & Zhou, M. (2021). The role of innovative approaches in aesthetic vocal performance. Musica Hodie, 21.
[4] Kulaha, T., & Segeda, N. (2021). Essential characteristics and content of the concept of contemporary pop vocal-performing thesaurus. KO KNOWLEDGE ORGANIZATION, 48(2), 140-151.
[5] Radolf, V., Laukkanen, A. M., Horáček, J., & Liu, D. (2014). Air-pressure, vocal fold vibration and acoustic characteristics of phonation during vocal exercising. Part 1: Measurement in vivo. Engineering Mechanics, 21(1), 53-59.
[6] Valenti, C. (2022). Vocal Music (Popular). In The Palgrave Encyclopedia of Victorian Women's Writing (pp. 1665-1670). Cham: Springer International Publishing.
[7] Anufrieva, N., Limanov, S., Mamonova, Y., Anchutina, N., & Kats, M. K. (2021). Teaching pop vocal performance to students: A comparison of methods used in Russia. Rast Müzikoloji Dergisi, 9(2), 2769-2782.
[8] Wang, Y. (2024). The effectiveness of innovative technologies to manage vocal training: The knowledge of breathing physiology and conscious control in singing. Education and Information Technologies, 29(6), 7303-7319.
[9] Chukwu, S. C., Egbumike, C. J., Ojukwu, C. P., Uchenwoke, C., Igwe, E. S., Ativie, N. R., ... & Uduonu, E. M. (2025). Effects of diaphragmatic breathing exercise on respiratory functions and vocal sustenance in apparently healthy vocalists. Journal of Voice, 39(2), 564-e37.
[10] Maiti, A. B., Dutta, M., Samaddar, S., Bhattacharya Samaddar, S., Dey, U., & Pradhan, R. (2024). Effects of the Change in Respiratory Kinematics on the Singing Voice: Re-Visiting the Benefits of Diaphragmatic-Abdominal Breathing in Professional Singers with Problems in Pitch. Ear, Nose & Throat Journal, 01455613241305984.
[11] Devi, K. R. (2021). Importance of proper breathing in singing. Performing Arts, 2(2), 67-71.
[12] Sokolova, Y. V. (2023). Breathing problem in music performance. Publishing House "Baltija Publishing".
[13] Innocenti, L., Nicolò, A., Massaroni, C., Minganti, C., Schena, E., & Sacchetti, M. (2022). How to investigate the effect of music on breathing during exercise: methodology and tools. Sensors, 22(6), 2351.
[14] Sliiden, T., Beck, S., & MacDonald, I. (2017). An evaluation of the breathing strategies and maximum phonation time in musical theater performers during controlled performance tasks. Journal of Voice, 31(2), 253-e1.
[15] Zeshan, J., & Thanawat, A. (2025). THE APPLICATION OF SPEECH LEVEL SINGING TRAINING SYSTEM IN CHINESE POP SINGING. Asia Pacific Journal of Religions and Cultures, 9(1), 147-158.
[16] Suryati, S. (2023). The Use of Growl Vocal Technique to Enhance Voice Quality in Singing Pop Songs. Resital: Jurnal Seni Pertunjukan, 24(2), 167-175.
[17] Bruder, C., & Larrouy-Maestri, P. (2023). Classical singers are also proficient in non-classical singing. Frontiers in Psychology, 14, 1215370.
[18] Hollien, H. (2014). Vocal fold dynamics for frequency change. Journal of Voice, 28(4), 395-405.
[19] Echternach, M., Döllinger, M., Sundberg, J., Traser, L., & Richter, B. (2013). Vocal fold vibrations at high soprano fundamental frequencies. The Journal of the Acoustical Society of America, 133(2), EL82-EL87.
[20] Wu, Y., & Chong, H. J. (2023). Affective responses to singing voice in different vocal registers and modes. The Journal of the Acoustical Society of Korea, 42(1), 75-82.
[21] Meireles, A., & Mixdorff, H. (2020). Voice quality in low and high registers in two different styles of singing. In Proc. Speech Prosody (Vol. 2020, pp. 601-605).
[22] Santos, S. S., Montagner, T., Bastilha, G. R., Frigo, L. F., & Cielo, C. A. (2019). Singing style, vocal habits, and general health of professional singers. International archives of otorhinolaryngology, 23(04), 445-450.
[23] Cameron, J. E., Duffy, M., & Glenwright, B. (2015). Singers take center stage! Personality traits and stereotypes of popular musicians. Psychology of Music, 43(6), 818-830.
[24] Hakanpää, T., Waaramaa, T., & Laukkanen, A. M. (2021). Comparing contemporary commercial and classical styles: emotion expression in singing. Journal of Voice, 35(4), 570-580.
[25] Juntunen, M. L., Arlin, E. P., & Liira, K. (2023, May). Expression in popular music singing as embodied and interpersonal. In Frontiers in Education (Vol. 8, p. 1092736). Frontiers Media SA.
[26] Hu, M. (2022). Features of singing in Chinese pop and traditional music: The influence of the music genre on vocal music. Musica Hodie, 22.
[27] Spreadborough, K. (2022). Emotional tones and emotional texts: A new approach to analyzing the voice in popular vocal song. Music Theory Online, 28(2).
[28] van Doorn, I. (2021). Breathing Technique for Jazz/Pop Singers. Journal of Singing, 78(2), 247-253.
[29] Yehui, T., & Meeson, P. (2023). THE APPLICATION OF MASK SINGING IN POPULAR VOCAL MUSIC. Asia Pacific Journal of Religions and Cultures, 7(2), 311-325.
[30] Xie, X. (2021). On the Similarities and Differences Between Bel Canto and Popular Singing. Frontiers, 2(1).
[31] Ayushi Prakash,Mayur Rahul,Sarvachan Verma,Sonu Kumar Jha & Vikash Yadav. (2025). Multilingual Speaker Recognition using Mel-frequency Cepstral Coefficients and Gaussian Mixture Model. Recent Advances in Electrical & Electronic Engineering,18(5),637-643.
[32] Sorin Muraru & Cătălina-Lucia Cocianu. (2024). Spoken Digit Recognition using the k-Nearest-Neighbor method and Mel Frequency Cepstral Coefficients. Informatica Economica,28(2),5-16.
[33] Kethireddy Rashmi,Kadiri Sudarsana Reddy & Gangashetty Suryakanth V.. (2022). Exploration of temporal dynamics of frequency domain linear prediction cepstral coefficients for dialect classification. Applied Acoustics,188,
[34] Yusnita Mohd Ali,Nor Fadzilah Mokhtar,Emilia Noorsal,Aida Zulia Zulhanip,Asmalia Zanal,Siti Zubaidah Md Saad & Nur Aliza Abdul Samad. (2020). Voice Command Intelligent System (VCIS) for Smart Home Application using Mel-frequency cepstral coefficients and linear prediction coefficients. Journal of Physics Conference Series,1535(1),012008.
[35] Suzhen Yuan,Xianli Li,Shuyin Xia,Xianrong Qing & Jermiah D. Deng. (2025). Quantum color image edge detection algorithm based on Sobel operator. Quantum Information Processing,24(7),195-195.