ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Empirical analysis of vocal fold vibration characteristics of modern popular vocal singing techniques

Zhiyong Li[1,*]

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

SUMMARY: *In this paper, the vibration process of the vocal folds of the examinee is sampled by laryngoscopy and used as sample data. Five major classes of filters are used to remove the background noise of the vocal fold image in separate cases to optimize the segmentation effect of the background and the vocal folds. In order to improve the extraction accuracy of vocal fold vibration contour points, the OTSU algorithm is used to set adaptive thresholds for different regions of the image, and binary segmentation is performed to segment the grayed-out vocal fold image and the background image. Based on the extracted vocal fold vibration data, we calculated and compared the normal and abnormal vocal fold vibration characteristics using modern popular vocal singing techniques. Inadequate mastery of modern popular vocal singing techniques will lead to 10 types of problems such as excessive guttural sound, which is reflected in the vibration characteristics of the vocal folds, such as complex and disordered waveforms, high and low frequencies, and reduced peaks, etc., and overall fails to show the beauty of popular vocal music.*

KEYWORDS: *OTSU algorithm; adaptive thresholding; binarization segmentation; vocal fold vibration characteristics; popular vocal singing techniques*

1 Introduction

Pop vocal music is continuing to garner widespread attention and love as a popular musical performing art. Judging from the tremendous success of pop music and its continuously growing audience base, pop vocal music has become a major force in today's music industry. At the same time, talent shows and variety shows focusing on pop vocal music continue to emerge, making pop vocal music a major player in the entertainment industry. However, it is not easy to stand out and realize the artistic value in the highly competitive field of pop vocal music. In pop vocal singing, singing technique becomes a crucial part, and different techniques not only cover different emotional expression and artistic performance, but also show different vocal vibration characteristics [1-3]. In the field of modern vocal music teaching and training, teachers and singers pay too little attention to the integration of physiological aspects and acoustic principles, and traditional empiricism leads to singing problems such as vocal cord damage, difficulty in meeting volume and range demands, and lack of sound quality stability [4-7]. Focusing on the vibration characteristics of the vocal folds of singers, revealing the correlation between their characteristics and vocal performance, promotes the singers, maintains the stability of the sound quality, meets the volume and range requirements of different styles of songs, and at the same time reduces the phenomenon of vocal fold damage.

Literature [8] studied the characteristics of vocal fold vibration in the transition stage of the

---

*lizhiyong2988318@126.com

https://doi.org/10.65102/is2026516

tenor voice area, there is a phenomenon of contraction of the laryngeal outer structure in the process of transition, and the opening coefficient of the vocal folds is larger in the mode of transition from the alto to the falsetto area, which indicates that the vibration of the vocal folds will present different characteristics according to the mode of transition. Literature [9] emphasized that vocal folds vibrate significantly in the low and middle frequencies in different types of singing, and in the high frequency range, the vocal folds vibration characteristics of singers differ, and the depth of vibration varies even for soprano singers. Literature [10] used laryngeal high-speed video endoscopy to analyze the dynamic characteristics of vocal fold vibration of singers, and the analysis showed that the fundamental frequency, vocal fold perturbation, and open coefficient of the waveform in the vocal fold region of female singers were higher than those of male singers, but the vibration of the vocal folds of male singers was more regular. Literature [11] with the help of acoustic analysis, laryngeal electrographic testing and video laryngeal fibrillation microscopy, it was found that most singers were able to produce sounds with anterior vibratory vocal fold patterns under the chest, falsetto, and whistling zones and were able to produce damping phenomena. Literature [12] analyzed the vocal fold vibrational characteristics in modern popular vocal performance techniques through the vocal fold vibrational components in dynamic vocal fold image sequences, pointing out that different singing techniques lead to differences in the vocal fold spectra, and that the vibrational frequencies of key vocal fold nodes in singing also show differences.

In this paper, videos of vocal cord vibrations of detectors singing using modern popular vocal singing techniques are captured to establish a vocal cord vibration characterization dataset. In the pre-processing stage of the dynamic video image, the background noise of the image is removed and the detailed features of the image are retained by a multi-class filter such as a mean filter. After that, a threshold segmentation method based on gray value features is chosen to segment the background and vocal folds into binarized images, and the adaptive thresholding method of OTSU algorithm is used to quickly distinguish different regions and improve the segmentation accuracy. The vibration contours of the vocal folds at the standard reference point are extracted, and the average rate and amplitude of vocal fold vibration are calculated when using modern popular vocal singing techniques to analyze the characteristics of vocal fold vibration.

2 Acquisition of vocal fold vibration images and feature extraction

2.1 Acquisition of vocal fold vibration samples

The subject is seated in a sitting position and is given surface anesthesia of the larynx, usually 2-3 times, until the patient's gag reflexes have largely disappeared. The test subject continuously made the “i” sound for more than 5 seconds, trying to keep the decibel around 85, the test subject exposed the laryngeal cavity through the rigid tube laryngoscope, and the entire vocal folds were exposed to ensure the completeness of the samples obtained. When the test subject finished the pronunciation, the laryngoscope was removed, and the complete vibration of the vocal folds was captured and saved as AVI lossless video. When the examinee's pronunciation is finished, the laryngoscope is removed, and the complete vocal fold vibration captured is saved in AVI lossless video format, and the video frame rate is set to 2500 fps. Considering that there may be occasional differences in the examinee's pronunciation due to emotional tension, such as out-of-tune pronunciation and off-key pronunciation, the same examinee needs to carry out 3 to 4 samples, and a better sample is selected for the extraction of the information. After obtaining the video information, the program was written to extract information about the vibration characteristics of the vocal folds at the 1/4 junction, 1/3 junction and 1/2 junction on both sides of the vocal folds. Figure 1 shows the process of acquiring the original image of vocal fold vibration.

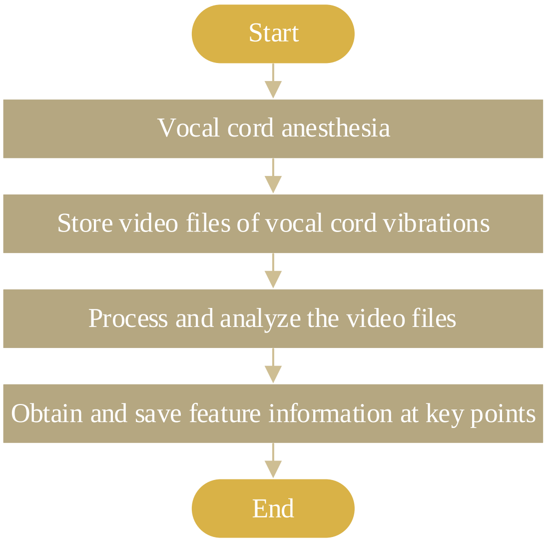


Figure 1: Original image acquisition process of vocal cord vibration

2.2 Preprocessing of dynamic vocal fold images

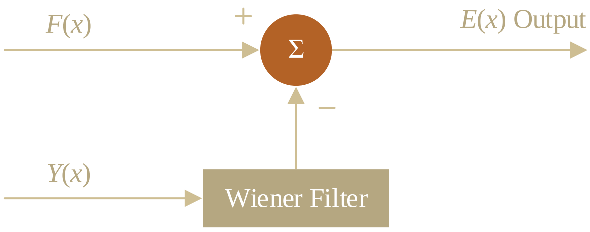
The pre-processing of dynamic images is to recognize each sorted dynamic digital image by means of a recognition module. The main role is to reduce the amount of irrelevant information in the image, highlight the useful information to be processed, enhance the detectability of the information in question and simplify the data to the greatest extent possible, so as to achieve improved feature extraction, dynamic image segmentation, matching and recognition, and to increase the reliability of dynamic image information recognition.

2.2.1 Background denoising of vocal tract images

Image denoising is the process of reducing noise in digital images. Since there are more other muscles and organs of the larynx in addition to the vocal folds in the original image captured for the experiment. In order to simplify the extraction process and at the same time accurately extract the feature information of the dynamic image of the vocal folds, it is necessary to perform background denoising on the vocal folds part to filter out the irrelevant organ parts in the image. Common denoising methods are as follows:

1) Mean value filter: For the particle noise in the image obtained by scanning, the mean value filter using the neighborhood averaging method is very suitable.

2) Adaptive Wiener filter: the Wiener filter is a linear filter based on the minimum mean square error criterion. Figure 2 shows the basic Wiener filter structure. The basic principle is that when there are two signals $F(x)$ and $Y(x)$ attached to the filter at the same time, the typical state is that $F(x)$ contains a component related to the $Y(x)$ signal as well as a component $E(x)$ that is not related to it, the Wiener filter can be used to obtain the optimal estimation of the component related to the $Y(x)$ in a $F(x)$, and thus can subtract that value to the $F(x)$ to subtract that value from $E(x)$.



$$\hat{n}_k = \sum_{i=0}^{N-1} w(i)Y(x-i)$$

Figure 2: Basic Wiener filter

3) Median filter: This filter is a nonlinear smoothing filter, whose main function is to change the pixels with large differences in the gray value of the surrounding pixels to values close to the values of the other pixels in its field, so as to achieve the effect of eliminating isolated noise and protecting the edges of the image.

The method of use is to use some structure of two-dimensional sliding template, the pixels in the template in accordance with the numerical size of the pixel value is sorted to generate monotonically ascending (or descending) two-dimensional data sequence. The 2D median filtering output results in:

$$g(x,y) = \mathrm{med}\left\{ f\left(x-k, y-l\right), \left(k, l \in W\right) \right\} \tag{1}$$

where $f(x,y)$ and $g(x,y)$ are the original image and the processed image, respectively. W is the 2D template.

4) Wavelet denoising: this method can retain most of the wavelet coefficients contained in the image information, and can effectively maintain the image details.

Its main features are: low entropy, accompanied by the sparse distribution of wavelet coefficients, the entropy value of the image obtained has been reduced; multi-resolution, the denoising method uses the means of multi-resolution, which can can be very good portrayal of the image edges, breakpoints, spikes and other signals such as the non-smooth characteristics of the signal; de-correlation, the wavelet transform can make the noise have a tendency to whiten after the transform, which can be de-correlation effect on the signal.

(5) Morphological noise filter: the method begins with the removal of noise on the background, which can be selected than the size of the noise of the structural elements of the matrix, and then the resulting image is processed to remove the noise from the image.

2.2.2 Threshold segmentation based on gray value features

Threshold segmentation method is applicable to images where the background and target belong to different gray level ranges respectively. The method is an image preprocessing process necessary before performing feature extraction, image analysis, and pattern recognition. Image thresholding is the process of classifying and categorizing the set of pixels of an image according to different gray levels so that there is a certain common property within each region and the neighboring regions differ in this property. And depending on the objective to be achieved, one or more thresholds can be chosen to be used.

In this paper, in the vocal cords image, the grayscale map after grayscaling of the vocal cords part is quite different from the image information after grayscaling of the background region, so in this paper, we choose to use the threshold segmentation based on the grayscale value features.

Thresholding method based on gray values is a common parallel region technique. The transformation of the input image f to the output image g of this thresholding method is as follows:

Where T is the threshold value set, for the target element region $g(i, j) = 1.0$ in the image and for the background element region $g(i, j) = 0.0$ in the image.

Determining a suitable threshold value is the key point of the algorithm, when the threshold value is selected appropriately, the background and target regions can be segmented perfectly. After selecting the threshold value T, the grayscale value of each pixel point is compared with the threshold value one by one to divide the original image into the target region and the background region, the segmentation of pixels can be realized in parallel processing, and the results obtained from the segmentation can directly give the results of each region.

For the case where a binary image needs to be obtained quickly and in real time, an automatic thresholding method can also be used.The OTSU algorithm is an efficient algorithm for obtaining a binarized image.The OTSU method works by dividing the original image into two parts, the foreground and the background, using a threshold value. The foreground part is represented by $n1$, $csum$, and $m1$ which indicate the number of points, mass moment, and average gray level of the foreground under the current threshold, and the background part is represented by $n2$, $sum-csum$, and $m2$ which indicate the number of points, mass moment, and average gray level of the background under the current threshold, respectively. The selected optimal threshold should satisfy the situation where the difference between the background and the foreground is the largest. The key point of this method is how to choose the appropriate standard to measure the difference, and the standard to measure the difference in the OTSU algorithm is chosen as the adaptive thresholding method to solve the problem. In some cases, the contrast between the target object and the background is not the same everywhere in the image, so it is difficult to completely separate the target and the background with the same threshold value. At this point, different thresholds are used to segment each local feature of the image separately. In the actual processing, we can first divide the image information into several sub-regions according to the goal to be achieved and then select the threshold value separately, or dynamically select the threshold value at each point according to the rules in a certain neighborhood range.

2.3 Acquisition of vocal fold vibration information

2.3.1 Algorithm design for point set extraction of vocal fold vibration contours

In addition to the vocal folds, there are more muscles and organs in the larynx in the original image collected from the experiment. In order to accurately extract the vocal fold contour and simplify the extraction process at the same time, so only the part containing the vibration of the vocal folds is processed and calculated.

The detection and extraction of the target contour point set is divided into the following steps:

1) In order to reduce the influence of reflections from the target surroundings and the shooting angle, the region of interest (ROI) containing the vocal folds in the image is manually determined from the color original image as an effective image sequence for subsequent processing.

2) After acquiring the desired ROI region, in order to improve the accuracy of target region information extraction and at the same time reduce the amount of raw data for subsequent calculation, the target image needs to be grayed out to reduce the amount of subsequent computational processing.

3) In order to extract the sequence image of the vibration contour of the vocal folds, a thresholding process is used to eliminate the influence of the surrounding environment and obtain a binarized image. The binarized sequence image of the target region takes θ as the binarization threshold:

$$B(x,y,t)=\begin{cases}255, & \left(F(x,y,t)>\theta\right)\\ 0, & \left(\mathrm{Other}\right)\end{cases} \tag{2}$$

4) In order to obtain the target contour, the image contour is obtained by means of image feature extraction, and only the outermost points of the binarized image of the vocal cords are kept and stored in the set. This point set is the extracted vocal fold contour points.

2.3.2 Calculation of average rate and amplitude of vocal fold vibration

The vibration contour of the vocal folds is elliptical, and the upper endpoint $A(x_1,y_1)$ and the lower endpoint $B(x_2,y_2)$ of the vocal folds are manually selected as the reference points $C_n(X_n,Y_n)$ based on the experience of clinical medicine. The coordinates of the x-axis and y-axis of each reference point are as follows:

$$X_n=\frac{x_1-x_2}{4}+x_1\left(n=1,2,3\right) \tag{3}$$

$$Y_n=\frac{y_1-y_2}{4}+y_2\left(n=1,2,3\right) \tag{4}$$

In order to get the vibration points of the reference point on both sides of the sound band, the coordinates of each point in the sound band contour point set $N(x,y)$ are substituted into the following equation:

$$a=\left(\left(x_1-X_n\right),\left(y_1-Y_n\right)\right) \tag{5}$$

$$b=\left(\left(x-X_n\right),\left(y-Y_n\right)\right) \tag{6}$$

$$\cos\alpha=\frac{a\cdot b}{|a|\cdot|b|} \tag{7}$$

Comparing the obtained $\cos\alpha$ values, the reference point C_n can be derived to correspond to the points $L_i(x_i,y_i)$ and $R_i(X_i,Y_i)$ which have the smallest $\cos\alpha$ values on the left and right sides of the vocal folds, respectively. The reference points correspond to the left side amplitude A_L and the right side amplitude A_R formulas are as follows:

$$A_L=\sqrt{\left(x_i-X_n\right)^2+\left(y_i-Y_n\right)^2} \tag{8}$$

$$A_R=\sqrt{\left(X_i-X_n\right)^2+\left(Y_i-Y_n\right)^2} \tag{9}$$

The reference point corresponds to the average rate of vibration of the left vocal fold V_L and the average rate of vibration of the right vocal fold V_R formulas are as follows:

$$V_L = \frac{\sum_{i=1}^{M}\left(\left|C_n D_i\right| - \left|C_n D_{i+1}\right|\right)}{t \cdot M} \tag{10}$$

$$V_R = \frac{\sum_{i=1}^{M}\left(\left|C_n E_i\right| - \left|C_n E_{i+1}\right|\right)}{t \cdot M} \tag{11}$$

where: M is the total number of captured video frames, and t is the time interval between two adjacent frames.

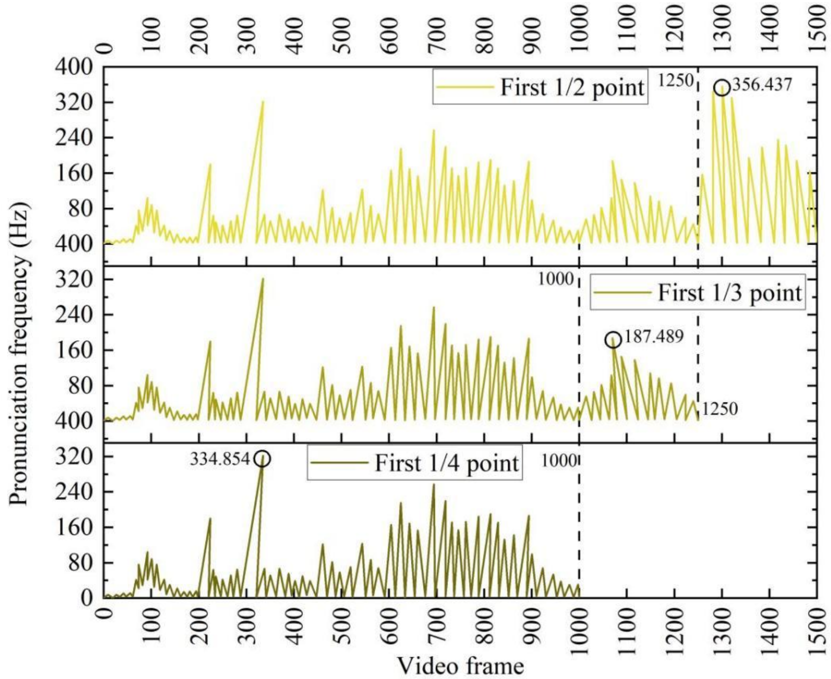
3 Example of vocal fold vibration characterization

3.1 Frequency and amplitude analysis of vocal fold vibration of collected samples

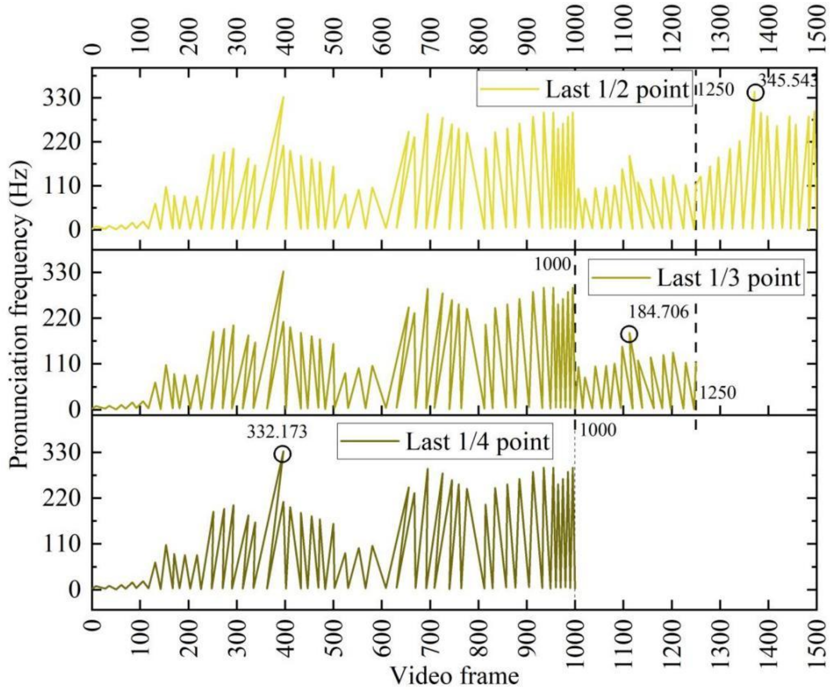
3.1.1 Vocal Sampling in Modern Popular Vocal Singing Techniques

The vocal fold vibration samples of 30 subjects were sampled when they sang using modern popular vocal singing techniques, and the vocal fold samples of the 30 subjects were preprocessed using filters, threshold segmentation algorithms, and so on. Afterwards, the vocal fold vibration contour points of the test subjects were extracted and the average rate and amplitude of their vibration were calculated.

Fig. 3 shows the sampling results of the vocal fold articulation of Detectee A at the first 1/4 point, the first 1/3 point, the first 1/2 point, the second 1/2 point, the second 1/3, and the second 1/4 point. From the sampling results, it can be seen that the vocal folds pronounced at 334.854 Hz for the first 1/4 points, 187.489 Hz for the first 1/3 points, 356.437 Hz for the first 1/2 points, 184.706 Hz for the second 1/3 points, and 332.173 Hz for the second 1/3 points.The pronunciation frequency of the reference points of the vocal folds in the second half of the segment was slightly lower than that of the The first half of the vocal fold reference point, this subject's vocal fold articulation is normal. From the sampling of the vocal folds of this participant, it can also be judged that the sampling effect is good.



(a) Reference point of vocal cords in the first half of the pronunciation frequency



(b) Reference point of vocal cords in latter part of the pronunciation frequency

Figure 3: Sampling results of reference points on vocal cords of tested individual

3.1.2 Analysis of average frequency and amplitude of reference points

After preprocessing the sampled dynamic images, the mean values of vibration frequency and amplitude values were calculated for the six vocal fold vibration reference points of the 30 subjects in this sampling. P-values were also calculated by paired t-test. Table 1 shows the average frequency and amplitude of the 6 vocal fold reference points of the 30 subjects. The p-values of the mean vibration frequency and amplitude values of the 3 reference points of the first half of the vocal folds and the 3 reference points of the second half of the vocal folds are greater than 0.05, so there is no statistically significant difference when analyzed in terms of the mean values.

Table 1: Average frequency and amplitude of the 6 vocal cord reference points

	Vocal cord reference points					
	First 1/4 point	First 1/3 point	First 1/2 point	Last 1/4 point	Last 1/3 point	Last 1/2 point
Frequency (Hz)	334.854 ± 66.828	321.293 ± 64.958	356.437 ± 79.948	331.173 ± 92.169	317.114 ± 111.485	345.543 ± 103.497
P-value	0.274			0.216		
Amplitude	30.485 ± 10.324	27.172 ± 11.438	29.396 ± 9.275	29.367 ± 9.751	26.458 ± 10.652	28.652 ± 10.864
P-value	0.394			0.320		

Analyzing only the average frequency and amplitude of the vocal fold reference points does not completely determine that the frequency and amplitude of the vocal fold vibration of the subjects are at a normal level, and therefore summarizes the frequency of vibration of the vocal folds of each of the 30 subjects in Table 2 and the amplitude of vibration of the vocal folds of each of the 30 subjects in Table 3.

Most of the 30 subjects' respective vocal fold vibration frequencies were in the average range, but some of the subjects had abnormal vocal fold vibration frequencies at one or more reference points. For example, both subjects 3 and 5 showed an abnormally high vocal fold vibration frequency at the position of the first 1/4 point, and the vocal fold vibration frequency of subject 5 even reached 784.353 Hz. The vocal fold vibration amplitude of each of the 30 subjects also showed a pattern similar to that of the vocal fold vibration frequency, with a higher number of abnormal amplitude cases. For example, the amplitude of all six vocal fold vibration reference points of examinee 8 were above 50 dB. There were more abnormal situations when the tested persons applied modern popular vocal singing techniques, and their vocal fold vibration characteristics need to be further specifically analyzed to study the relevant laws.

Table 2: Individual vocal vibration frequencies of the 30 tested individuals ()

Tested individuals	First 1/4 point (Hz)	First 1/3 point (Hz)	First 1/2 point (Hz)	Last 1/4 point (Hz)	Last 1/3 point (Hz)	Last 1/2 point (Hz)
1	393.736	386.251	383.291	239.004	396.232	242.046
2	279.295	256.335	294.853	277.206	281.791	298.054
3	**501.682**	365.604	436.385	410.333	404.178	439.586
4	432.794	377.125	422.352	404.701	412.259	425.553
5	**784.353**	381.352	276.489	376.262	386.849	279.659
6	393.737	390.735	**583.296**	385.641	396.233	386.497
7	286.542	273.541	266.105	368.452	**589.038**	269.306
8	417.175	314.176	398.917	401.263	419.671	402.118
9	378.104	375.101	367.662	490.016	380.246	370.863
10	409.359	306.358	398.911	401.262	411.855	402.112
11	269.487	274.494	297.047	229.393	271.983	**600.248**
12	370.295	367.293	359.854	362.266	372.791	363.055
13	370.294	367.296	309.241	362.201	352.759	312.442
14	432.797	329.791	402.356	284.705	428.599	449.040
15	354.676	351.675	**644.237**	346.587	357.172	347.438
16	268.026	272.291	334.852	237.202	270.522	338.053
17	346.851	343.858	346.413	338.763	349.347	349.614
18	378.103	375.102	367.664	370.018	380.599	370.865
19	385.928	**772.926**	375.488	397.839	388.424	378.689
20	370.296	367.298	359.853	423.342	372.792	363.054
21	402.744	308.407	399.969	378.321	405.244	403.187
22	433.852	276.543	298.105	**587.467**	414.348	301.306
23	**685.411**	369.342	360.912	370.511	**687.907**	364.113
24	394.795	369.345	310.299	403.322	397.291	313.655
25	287.466	331.284	403.414	382.075	289.962	406.615
26	418.233	353.724	345.295	413.543	420.729	348.496
27	379.162	274.324	335.932	231.452	**581.658**	269.133
28	410.417	345.907	347.471	364.325	412.913	350.672
29	386.986	377.151	368.722	364.216	389.482	371.923
30	**571.354**	374.975	376.546	286.764	205.629	379.747

Table 3: Amplitude of vocal cord vibrations for each of the 30 tested individuals

Tested individuals	First 1/4 point (dB)	First 1/3 point (dB)	First 1/2 point (dB)	Last 1/4 point (dB)	Last 1/3 point (dB)	Last 1/2 point (dB)
1	21.445	15.734	24.936	20.313	25.992	22.696
2	18.844	21.149	22.335	25.745	23.361	28.128
3	42.548	44.853	38.671	49.449	39.727	51.832
4	44.242	46.547	47.733	51.143	48.789	53.526
5	21.044	23.349	24.535	27.945	25.591	30.328
6	44.146	46.451	47.637	51.047	48.693	53.435
7	9.348	38.610	12.839	43.206	13.895	45.589
8	**51.547**	**53.852**	**55.038**	**58.448**	**56.094**	**60.831**
9	25.348	27.653	28.839	32.249	29.895	34.632
10	10.441	12.746	13.932	17.342	14.988	19.725
11	24.144	26.449	27.635	31.045	28.691	33.428
12	27.048	29.353	30.539	33.949	31.595	36.332
13	26.847	29.152	30.338	33.748	31.394	36.131
14	39.146	41.451	42.637	46.047	43.693	48.437
15	24.345	26.265	27.836	30.861	28.892	33.244
16	14.442	16.747	17.933	21.343	18.989	23.726
17	37.543	39.848	41.034	44.444	42.059	46.827
18	36.049	38.354	39.354	42.965	40.641	45.348
19	29.541	31.846	33.032	36.442	34.088	38.825
20	29.545	31.845	33.036	36.441	34.092	38.824
21	11.682	13.987	15.173	18.583	16.229	20.966
22	25.383	27.698	28.874	32.294	29.923	34.677
23	28.286	30.591	31.777	35.187	32.833	37.579
24	28.087	30.392	31.578	34.988	32.634	37.371
25	40.382	42.687	43.873	47.283	44.929	49.666
26	25.588	27.593	29.079	32.189	30.135	34.572
27	15.683	17.982	20.121	22.578	21.177	24.961
28	38.781	41.086	42.272	45.682	43.328	48.065
29	40.809	43.114	44.537	47.771	45.593	50.154
30	20.161	22.465	23.652	27.061	24.708	29.431

3.2 Analysis of abnormal vocal fold vibration problems

Comparing the vocal folds vibration of 30 subjects when applying modern popular vocal singing techniques with the standard vocal folds vibration of senior singers when applying modern popular vocal singing techniques, and analyzing the reasons for the abnormalities of vocal folds vibration of the subjects. Table 4 shows the abnormalities of vocal fold vibration and the characteristics of the test subjects. 30 test subjects had 10 abnormalities of vocal fold articulation, such as excessive laryngeal sound, nasal sound, leakage of air, breath holding, voice squeezing, lack of resonance, high dentition, unclear biting, unstable breath control, incomplete vocal fold vibration, etc. The 10 conditions were manifested in the frequency and amplitude of the vocal folds, such as the lack of low-frequency, and the increase in the middle-low-frequency loudness, etc. The results are summarized in the following table.

Based on the segmentation of the background and vocal folds of the vocal fold vibration image by gray value features, the contour points of the vocal fold vibration are extracted, the average rate and amplitude of the vocal fold vibration are calculated, and then compared with the standard vocal fold vibration of the senior singers, the abnormalities of the vibration of the vocal folds among the different examinees can be identified more accurately.

Table 4: Abnormalities and characteristics of vocal cord vibration

Abnormal vocal cord vibration	Problem characteristics
Excessive laryngeal sound	Low-frequency deficiency, increased mid-low frequency loudness
Nasal sound	Mid-high frequency position deficiency
Air leakage	Mid-high frequency deficiency, high-frequency enhancement
Breath holding	Mid-range area deficiency
Squeezing the voice	Mid-range area deficiency with overall lower loudness
Lack of resonance	High-frequency to mid-range area loudness is too high
Excessive dental sounds	Small range increase in high-frequency part
Unclear articulation	Small range decrease in mid-high frequency part
Unstable breath control	Large range decrease in high-frequency part
Incomplete vocal cord vibration	Low-medium-high frequency distinction is not good

3.3 Comparison of vocal fold vibration characteristics of modern popular vocal singing techniques

3.3.1 Comparison of the vibration spectra of excessive and normal laryngeal sounds

Figure 4 compares the vocal fold vibration spectra of the examinee with excessive laryngeal tone with those of the advanced vocalist with normal laryngeal tone. When applying modern popular vocal singing techniques, the vocal folds of the normal laryngeal voice were in a more relaxed closed state, the waveform movement was more organized, the 13 peaks were prominent, and the sound quality was clear. On the other hand, the waveform movement of the vocal folds with excessive laryngeal tone has frequent ups and downs, and is obviously in a disordered state, with the peaks and valleys not prominent, and the sound quality has more noises, and the purity of the timbre is not prominent.

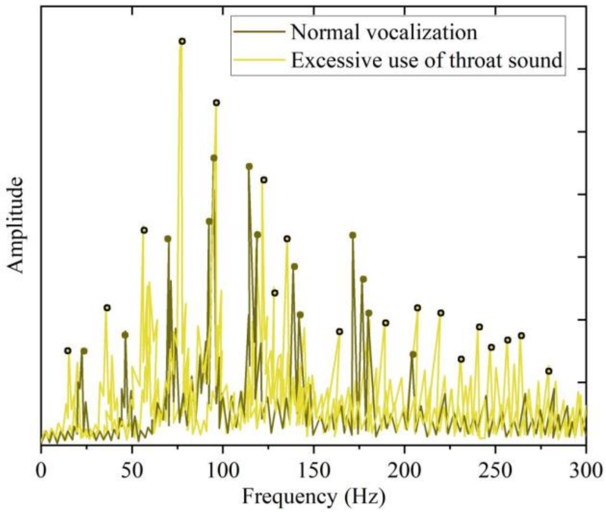


Figure 4: Comparison of heavy throat sounds and normal vibration spectra

3.3.2 Comparison of the spectra of full and localized vibration of the vocal folds

Figure 5 shows a comparison of the vocal fold vibration spectra of the examinee with fully vibrated vocal folds and that of the advanced vocalist with partially vibrated vocal folds. Comparison of the spectrograms shows that when the vocal folds are under full vibration, the waveform is more complex, with fewer peaks, only about 7, and the overtone columns are more acute, thus reflecting a more tense vocal state. The harmonic overtones are more, and the waveform maintenance degree is lower than that in the partial vibration state of the vocal folds, and it is basically impossible to maintain the waveform at about 150 Hz. The recognition of the voice is poor. In addition, compared to the spectrum of the vocal folds under localized vibration, there are more “burrs” in the waveform under full vibration, and the stability of the sound is lower, with a sharper timbre.

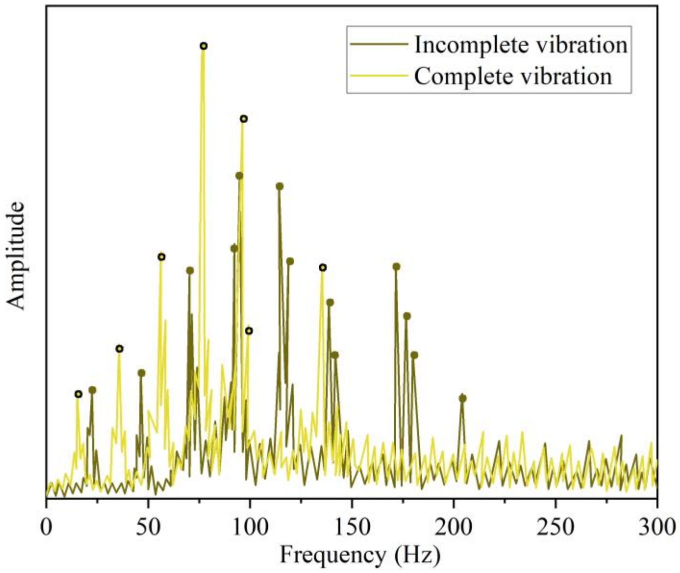


Figure 5: Comparison of spectra of complete and local vocal cord vibration

3.3.3 Spectral comparison of stable and unstable breath-holding

Figure 6 shows a comparison of the vocal fold vibration spectra of the participant with unstable breath control and that of the senior vocalist with stable breath control. Comparing the spectra of the two different states, it can be seen that the waveform peaks in the spectrum of the unstable breath control state are significantly reduced, with only about 7 peaks. When the breath control is stable, the overtones are spaced at small intervals, with close spacing values and similar overtones following each spacing value. In the unstable breath control condition, the overtone columns only show obvious responses in the intrinsic frequency band, which reflects that in the first half of the breath control unstable singing, the gene tone is prominent, with obvious texture, but in the second half of the test, the breath sound is louder than the gene volume, which affects the purity of the tone, and at the same time, the overall volume is also decreased.

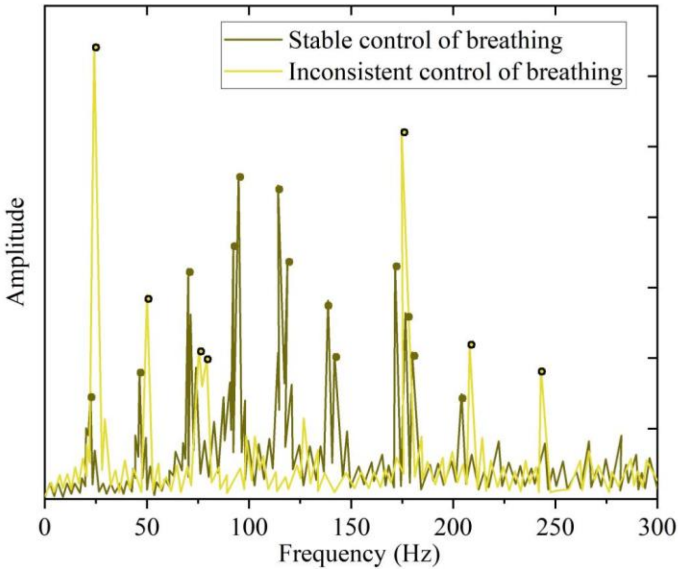


Figure 6: Comparison of the spectra of stable and unstable breath control

4 Conclusion

In this paper, the adaptive threshold of OTSU algorithm is utilized to efficiently segment the background and vocal folds of the vocal fold vibration image, which improves the extraction effect of the vocal fold vibration contour points, and also makes the analysis of the vocal fold vibration characteristics more accurate.The calculation results of the frequency and amplitude of the 6 vocal fold vibration reference points proved that the vocal fold vibration contour points have a high accuracy. The mining of 10 types of abnormal vocal fold vibration problems and the comparison of 3 types of abnormal vocal fold waveforms also verified that the degree of mastery of modern popular vocal singing techniques would affect the final singing effect. When there is excessive laryngeal tone, complete vibration of the vocal folds, unstable breath control, etc., the vibration of the vocal folds will be disorganized, and the waveforms will be high, low and disappearing, etc., which can not restore the beautiful sound quality of pop vocal music.

References

[1] Caffier, P. P., Nasr, A. I., Rendon, M. D. M. R., Wienhausen, S., Forbes, E., Seidner, W., & Nawka, T. (2018). Common vocal effects and partial glottal vibration in professional nonclassical singers. Journal of Voice, 32(3), 340-346.

[2] Elbarougy, R. (2019). Acoustic analysis for chest-to-head register transition in singing voice. International Journal of Computer Applications, 177, 11-16.

[3] Wang, D. (2022). Emotional expression in vocal skill from the perspective of multiculturalism. Música Hodie, 22.

[4] Fantini, M., Fussi, F., Crosetti, E., & Succo, G. (2017). Estill Voice Training and voice quality control in contemporary commercial singing: an exploratory study. Logopedics Phoniatrics Vocology, 42(4), 146-152.

[5] Kwok, M., & Eslick, G. D. (2019). The impact of vocal and laryngeal pathologies among

professional singers: a meta-analysis. Journal of Voice, 33(1), 58-65.

[6] Folomieieva, N. A. (2020). MASTERING VOCAL-PERFORMING TECHNIQUES BY STUDENTS IN A POP SINGING CLASS. Publishing House “Baltija Publishing”.

[7] ERDEM, D. A. (2021). VOICE PHYSIOLOGY AND VOCAL TRAINING. Asst. Prof. Başak DANACI POLAT, 171.

[8] Echternach, M., Burk, F., Köberlein, M., Herbst, C. T., Döllinger, M., Burdumy, M., & Richter, B. (2017). Oscillatory characteristics of the vocal folds across the tenor passaggio. Journal of Voice, 31(3), 381-e5.

[9] Song, U. J., Ahn, I. S., Kim, M. S., & Bae, M. J. (2018, June). A Study on the Stability of Ultra-High Frequency Vocalization of Soprano Singers. In International Conference on Software Engineering Research, Management and Applications (pp. 245-253). Cham: Springer International Publishing.

[10] Khairuddin, K. A. M., Ahmad, K., Proehoeman, S. C., Ibrahim, H. M., & Yan, Y. (2024). Preliminary findings of vocal fold vibratory characteristics of singers analyzed by laryngeal high-speed videoendoscopy. Journal of Voice.

[11] Fantini, M. (2024). The Physiology of Vocal Damping: Historical Overview and Descriptive Insights in Professional Singers. Journal of Voice.

[12] Li, Z. (2025). Analysis of Vocal Cord Vibration Characteristics in Modern Popular Vocal Performance Techniques. International Journal for Housing Science & Its Applications, 47(1).