International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 2909-2919

https://doi.org/10.70517/ijhsa464242

A Study on the Resonance Cavity Regulation Mechanism of Bel Canto Singing Based on Acoustic Analysis

Guili Yao[1,*]
[1] School of Arts (Weihai), Shandong University, Weihai, Shandong, 264200, China
Corresponding authors: (e-mail: lsxieping@163.com).

Abstract This paper employs time-frequency transformation technology and the Transformer self-attention mechanism to construct a timbre detection model, combined with a source-filter model to perform quantitative analysis of the formant parameters of five vowel categories. The experiment uses the Hann window-optimized STFT algorithm to extract time-frequency features and employs a multi-head self-attention mechanism to model acoustic feature correlations. The study found that the vocal range of bel canto singing in songs is slightly larger than that of ethnic singing. In the vowels /a/, /u/, and /o/, the F1 and F2 values of incorrect vocalization in the bel canto singing style are smaller than those of standard vocalization. In the vowels /i/ and /e/, the F1 values of oral cavity group vocalization are larger than those of the head cavity group, while the F1 values of standard vocalization are between the two groups. Conversely, in F2 values, the oral group is smaller than the head cavity group, but the standard vocalization F2 values still fall between the two groups.

Index Terms Bel Canto Singing Technique, Time-Frequency Transformation Technology, Transformer, Source-Filter Model, Formant, Resonance Cavity

I. Introduction

Bel canto singing originated in Europe during the Renaissance. It enables free and rapid transitions between different vocal sounds and better integrates the artistic content contained in these sounds, thereby providing listeners with a superior aesthetic experience [1]-[3]. The artistic expression of bel canto singing is based on a sophisticated control of the human vocal apparatus and a deep understanding of acoustic principles. This singing technique is far from merely pursuing volume expansion; instead, it establishes a precise and coordinated physiological operation and acoustic resonance mechanism, ultimately presenting a vocal art form that combines penetrating power, tonal layers, endurance, clear diction, and emotional depth [4]-[6]. Resonance is an indispensable and important element in bel canto singing, directly influencing the artistic effect of the performance.

The fundamental frequency produced by the vocal cords is weak, but when resonance is utilised, the volume becomes loud. This is because when resonance develops, it often requires the simultaneous use of different resonance chambers, such as the larynx, pharynx, oral cavity, nasal cavity, or the entire cranial cavity and thoracic cavity, all of which are resonance chambers. These can be developed through training to function as a unified whole, Overall resonance enables the volume to be continuously amplified, resulting in a loud and beautiful tone [7]-[11]. In vocal performance, the original sound produced by vocal fold vibration undergoes significant volume enhancement and tonal changes after resonance. Acoustic research indicates that the resonance process involves highly complex acoustic-physical changes, with resonance amplifying certain overtones while suppressing others [12]-[15]. The human body's resonant cavities form a unique vocal tract transmission system. Due to the elastic properties of the resonant cavity walls and their sound wave reflection characteristics, sound waves are continuously amplified as they propagate [16], [17]. The resonance process is also closely related to the volume and shape of the resonance cavity. Changes in the cavity directly affect the propagation direction and energy distribution of sound waves. The degree of matching between the fundamental frequency produced by vocal fold vibration and the resonance frequency generated by the resonance cavity determines the quality of the resonance effect [18]-[20]. A deep understanding of the acoustic principles of resonance is of great significance for adjusting the vocal position and controlling timbre in bel canto singing.

Effective resonance coordination directly affects vocal expressiveness and artistic appeal. accurate resonance positioning enables the voice to achieve optimal acoustic effects, enhancing its penetrating power and extension, resulting in good loudness and timbre. When singing in the high register, proper use of head resonance helps reduce vocal cord strain, yielding a bright and transparent sound quality. In the mid-to-low register, appropriately increasing chest resonance can make the voice more robust and dynamic [21]-[24]. Accurate control of resonance also helps

singers achieve natural transitions between different registers, maintaining tonal consistency and continuity. Scientific use of resonance not only enhances vocal expressiveness but also reduces fatigue in the vocal organs, promoting long-term vocal development. By precisely adjusting resonance, singers can demonstrate vocal diversity across different dynamics and emotional expressions, making performances more artistically compelling and emotionally impactful [25]-[29]. Therefore, the quality of resonance cavity regulation directly determines the overall artistic level of bel canto singing. Utilising acoustic analysis to reveal the resonance cavity regulation mechanisms in bel canto singing provides references for bel canto singing instruction, music composition, and vocal technique practice.

This paper first establishes a detection system based on time-frequency analysis, employing an STFT algorithm optimized with a Hann window to address the issue of time-domain information loss. A multi-head self-attention mechanism is introduced, leveraging the Transformer network for feature extraction. By integrating various algorithms such as spectral envelope estimation vocoders, discrete all-pole models, and cepstral coefficients, the system analyzes resonance states based on a source-filter model. Acoustic parameters of vocal sounds are measured for different singing techniques, and tonal differences between these techniques are evaluated. Through a systematic analysis of standard and error vocal samples, the system examines the regulatory mechanisms of resonance cavities and their acoustic characteristics in bel canto singing.

II. Design of tone detection and resonance state analysis technology for bel canto singing

Resonance regulation in vocal art is a core element determining the quality of tone color, especially in bel canto singing, where the coordinated operation of resonance cavities directly affects vowel recognition and acoustic performance. Existing research has mostly focused on the acoustic characteristics of single resonance cavities, lacking quantitative analysis of the mechanisms of multi-cavity interaction. This paper uses multimodal acoustic analysis technology to systematically study the dynamic regulation patterns of resonance cavities in bel canto singing.

II. A.Acoustic Characteristics Analysis

In the process of analyzing acoustic characteristics, frequency domain analysis is first performed. Fourier transform is used to convert the sound signal from the time domain to the frequency domain, and its spectral composition is analyzed in detail. The key to this step is to identify the amplitude and phase of each major frequency component in order to evaluate the sound quality of the sound source and possible noise pollution issues. Acoustic characteristic analysis also includes harmonic analysis, which can identify the nonlinear characteristics of the sound source. The principle of acoustic characteristic analysis technology is shown in Figure 1.

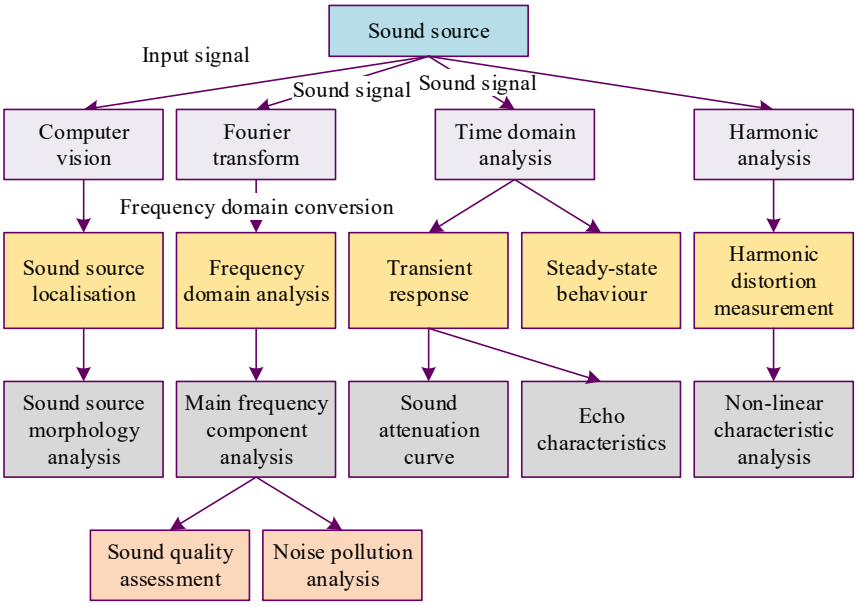

Figure 1: Principle of Acoustic Characteristic Analysis Technology

II. B. Tone detection technology

II. B. 1) Time-frequency transformation technology

Sound is generated by mechanical vibrations, and the original sound signal is time-domain information where the amplitude y varies with time t. Singing sound signals can be regarded as periodic signals over short periods of time, and the main characteristics of periodic signals can only be clearly manifested in the frequency domain. The audio signals collected by actual computing devices are discrete samples of continuous sound signals. Generally, the discrete Fourier transform (DFT) is used to transform the discrete sound signals in the time domain to the frequency domain. Let a segment of audio signal be $x[n](n=1,2,...,N)$ is a segment of audio signal, where N denotes the length of the signal, then the corresponding Fourier transform result $X[k]$ is expressed as:

$$X[k]=\sum_{n=0}^{N-1} e^{-\frac{2\pi}{N}nk} x[n],\ k=0,1,...,N-1 \tag{1}$$

Here, e denotes the natural logarithm base. From the above equation, we can see that the algorithmic complexity of DFT is $O(N^2)$, which incurs enormous computational overhead when processing high-sampling-rate audio. To address this issue, researchers employed the butterfly algorithm to eliminate redundant computations in the original DFT, resulting in the Fast Fourier Transform (FFT), whose complexity is reduced to $O(N\log N)$. After sound signals are transformed into the frequency domain via FFT, their temporal domain information becomes inaccessible, necessitating the use of the Short-Time Fourier Transform (STFT) to resolve this issue.

The essence of the STFT is the FFT transformation of a signal after “framing” and “window application.” “Framing” refers to dividing the original signal into several equal-length audio frames, with overlapping sections retained between adjacent frames to ensure smooth transitions. “Windowing” involves applying a window function to the audio frames to reduce spectral leakage. The window used in this paper is the Hanning window, whose formula is as follows:

$$Hanning(l)=0.5\left(1-\cos\left(\frac{2\pi l}{L-1}\right)\right) \tag{2}$$

In this context, L denotes the window length. After applying the STFT transformation to the signal, we can not only observe the frequency-domain information within each audio frame but also examine how these frames evolve over time. The transformation results collectively form a joint time-frequency distribution. The spectrogram effectively balances time-domain and frequency-domain information, clearly revealing the harmonic series distribution of this sound segment.

II. B. 2) Self-attention mechanism in Transformer

The Transformer was initially used as a backbone network in the field of natural language processing. In recent years, with continuous improvements to the Transformer, networks based on the Transformer have emerged as leaders among networks that were previously dominated by CNNs and RNNs, demonstrating excellent recognition performance in fields such as image and audio recognition. The principle of the self-attention mechanism in the Transformer is shown in Figure 2. This section will explain the self-attention mechanism and multi-head attention mechanism used in the Transformer model.

(1) Self-attention mechanism

The self-attention mechanism is the core module used by Transformer to discover global correlation features in data. The Transformer takes a sequence of feature vectors $X\in R^{n_x\times d_x}$ as input, where n_x is the length of the sequence and d_x is the dimension of the feature vector. Before the input X enters the self-attention mechanism module shown in Figure 2 (a), it first passes through three different projection matrices $W^Q\in R^{d_x\times d^k}$, $W^K\in R^{d_x\times d^k}$, and $W^V\in R^{d_x\times d^k}$, respectively, to three corresponding feature vector sequences: $Q(Query)$, $K(Key)$ and $V(Value)$ respectively, where d^k is the dimension of the input feature vector after projection. The process is shown below:

$$Q=XW^Q, K=XW^K, V=XW^V \tag{3}$$

The self-attention mechanism of the Transformer belongs to a key-value pair attention mechanism. Its core approach involves calculating the key Key in the key-value pair using the query value $Query$, and then retrieving the corresponding value $Value$ using Key. When the sources of $Query$, Key, and $Value$ are all from the same input, this attention mechanism is referred to as a self-attention mechanism. After transforming the input X into the corresponding $Query$, Key, and $Value$, the self-attention mechanism module of the Transformer calculates the corresponding attention output using the following formula:

$$Attention(Q,K,V) = Soft\max\left(\frac{QK^T}{\sqrt{d^k}}\right)V \tag{4}$$

Among them, d^k is the dimension of the feature vector. In equation (4), the transpose of Q is multiplied by K to perform a dot product operation between each feature vector in Query and all feature vectors in Key, thereby obtaining attention weights with global relevance. The attention weights are then scaled by $\sqrt{d^k}$ to ensure gradient stability during subsequent training, normalized using the Softmax function, and finally multiplied by the Value vector to obtain the output vector of this module.

$$Soft\max(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{d^k} e^{z_j}} \tag{5}$$

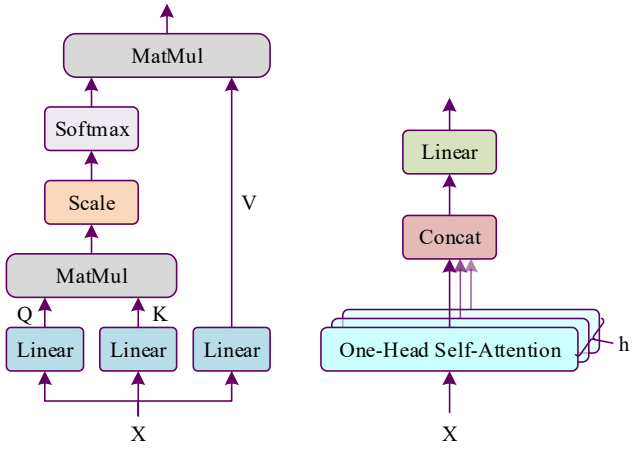

Figure 2: The Principle of Self-Attention Mechanism in Transformer

(2) Multi-head attention mechanism

The ordinary self-attention mechanism is a single-head self-attention mechanism, which uses only one projection matrix for each process of converting the input X into $Query$, Key, and $Value$. The feature subspace constructed by the single-head attention mechanism is very limited, making it difficult to achieve complex modeling. Therefore, Transformer uses a multi-head self-attention mechanism (MHSA) to improve the performance of its self-attention mechanism module. The structure of the multi-head attention mechanism is shown in Figure 2 (b), where multiple independent single-head self-attention operations are performed on the input X. A projection matrix used in a single-head self-attention mechanism is referred to as an attention head. Multiple attention heads are used to map the input X to multiple feature subspaces, and the corresponding single-head self-attention mechanism operations are performed within each feature subspace. Finally, the outputs of all attention heads are fused together to form the final output of the module. Assuming the number of attention heads is h, this process can be expressed as follows:

$$q_i = XW^{q_i}, k_i = XW^{k_i}, v_i = XW^{v_i}, i = 1...h \tag{6}$$

$$z_i = Attention(q_i, k_i, v_i), i = 1...h \tag{7}$$

$$MultiHead(Q,K,V) = Concat(z_1, z_2, ..., z_h)W^0 \tag{8}$$

Equations (6) and (7) describe the process of a single-head attention mechanism, where i is the index of the attention head, $W^{q_i} \in R^{d_x \times d_{head}}$, $W^{k_i} \in R^{d_x \times d_{head}}$, $W^{v_i} \in R^{d_x \times d_{head}}$ are three distinct projection matrices, and d_{head} denotes the dimension of the feature vector output from the projections within an attention head. It is evident that the computational processes of each attention head are independent of one another, with each attention head possessing an independent feature space. This enhances the diversity of feature subspaces while ensuring the proper functioning of the single-head self-attention mechanism. In equation (8), the output vectors from multiple

attention heads are concatenated and mapped through the output projection matrix $W^0 \in R^{hd_{head} \times d_k}$ to obtain the final output feature vector.

II. C.Resonance state analysis based on the source-filter model

The spectral envelope referred to in this paper refers to the frequency response of the filter in the source-filter model. The spectral envelope contains semantic and timbre information, and its accurate extraction lays the foundation for recognition, synthesis, and conversion. This section mainly introduces four algorithms related to spectral envelope extraction: vocoder, linear predictive coefficients, discrete all-pole, and cepstral coefficients.

II. C. 1) Spectral Envelope Estimation Vocoder

The method of estimating the spectral envelope using harmonics—Spectral Envelope Estimation Vocoder (SEEVOC)—finds the positions of each harmonic on the logarithmic amplitude spectrum based on the average fundamental frequency of the signal, retains the frequency and amplitude of the harmonics, and performs linear interpolation or third-order spline interpolation according to the original frequency axis to obtain the spectral envelope.

II. C. 2) Linear prediction coefficients

The basic idea behind linear predictive coefficients (LPC) is to approximate the current sample value using a linear combination of several past sample values. Linear prediction is based on a fully polar model and obtains linear predictive coefficients through minimum mean square error. In a fully-pole model, the signal s_n can be modeled as a linear combination of past-time sample values s_{n-k} and the current-time input u_n, as shown in Equation (9):

$$s_n = -\sum_{k=1}^{p} a_k s_{n-k} + G u_n \tag{9}$$

In the equation, G is the gain constant, a_k is the linear prediction coefficient, and p is the order of the model. The transfer function of formula (9) is shown in formula (10):

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{10}$$

In practice, however, the input u_n is unknown, so s_n can only be predicted using a linear combination of s_{n-k}, as shown in formula (11):

$$\tilde{s}_n = -\sum_{k=1}^{p} a_k s_{n-k} \tag{11}$$

Then, the prediction error e_n is shown in formula (12):

$$e_n = s_n - \tilde{s}_n = s_n + \sum_{k=1}^{p} a_k s_{n-k} \tag{12}$$

To solve for the linear prediction coefficients a_k, we can first take the derivative of the sum of squared errors, i.e., $\frac{\partial E}{\partial a_i} = 0$, $1 \le i \le p$, where $E = \sum e_n^2$. The commonly used autocorrelation method generally takes $-\infty < n < +\infty$. The solution to the derivative of the mean squared error is shown in formulas (13) and (14):

$$\sum_{k=1}^{p} a_k R(i-k) = -R(i), 1 \le i \le p \tag{13}$$

$$E_p = R(0) + \sum_{k=1}^{p} a_k R(k) \tag{14}$$

Among them, $R(k)$ is the autocorrelation of the signal. According to the Levinson-Dubin recursive algorithm, the solution a_k of formula (13) can be calculated with relatively little computation. According to $G^2 = E_p$, the gain G can be calculated. Substituting G and a_k into formula (10) yields the frequency response of the filter, i.e., the spectral envelope.

II. C. 3) Discrete poles

To overcome the shortcomings of linear prediction in extracting high-frequency signals, the discrete all-pole (DAP) model was proposed. Its main idea is to use the $\hat{P}(\omega_m)$ obtained from linear prediction modeling to approximate the discrete amplitude spectrum $P(\omega_m)$ and minimize the error. The error criterion is shown in formula (15):

$$E_{IS} = \frac{1}{N}\sum_{m=1}^{N}\left[\frac{P(\omega_m)}{\hat{P}(\omega_m)} - \ln\frac{P(\omega_m)}{\hat{P}(\omega_m)} - 1\right] \tag{15}$$

where N denotes the number of discrete points, and the definition of $\hat{P}(\omega_m)$ is shown in formula (16):

$$\hat{P}(\omega) = \left|H(\omega)\right|^2 = \frac{1}{\left|A(\omega)\right|^2} = \frac{1}{\left|\sum_{k=0}^{p} a_k e^{-jwk}\right|^2} \tag{16}$$

Substituting equation (16) into (15) and performing minimization, we obtain the prediction coefficients a_k of the discrete all-pole model.

II. C. 4) Inverse coefficient

Spectral inversion is widely used in various fields of signal processing and is mainly divided into complex spectral inversion and real spectral inversion. Complex spectral inversion is obtained by taking the logarithm of the spectrum and then performing the inverse Fourier transform, while real spectral inversion is obtained by taking the modulus of the spectrum and then taking the logarithm. This paper uses real spectral inversion, the expression of which is shown in formula (17):

$$C(l) = \sum_{k=0}^{K} \ln\left(\left|X(k)\right|\right) e^{i\frac{2\pi kl}{K}} \tag{17}$$

Here, $X(k)$ denotes the DFT spectrum of the signal, K denotes the number of discrete points, and l denotes the l th dimension of the cepstral coefficients.

In the source-filter model, the time-domain signal of human voice is the convolution of glottal excitation and vocal tract impulse response. When transformed into the frequency domain, it becomes a product form. Taking the logarithm converts the product term into a sum form, meaning that the cepstral spectrum of the human voice signal is the sum of the cepstral spectrum of glottal excitation and the cepstral spectrum of vocal tract impulse response. The glottal excitation cepstrum is distributed at integer multiples of the fundamental period, while the cepstrum of the vocal tract impulse response is primarily distributed within the range between the origin and one fundamental period. Therefore, the initial low-frequency portion of the cepstrum of the vocal signal primarily contains information about the vocal tract transfer function, while the subsequent high-frequency portion primarily reflects information about the fundamental frequency. Thus, by taking the first few dimensions of the cepstrum coefficients and performing a Fourier transform, an estimate of the spectral envelope can be obtained.

III. Analysis of resonance cavity adjustment in bel canto singing based on acoustic analysis

In bel canto singing, the vocal technique typically involves a blend of head voice and chest voice, with a mixed resonance (nasal resonance, oral resonance, and head resonance). By adjusting the balance of resonance centers through mixed voice, an overall resonance effect is achieved, resulting in a smooth and full-bodied sound. The underlying reason is that it is based on the linguistic system represented by Italian, which belongs to the Indo-European language family. Therefore, Italian pronunciation has characteristics such as a lower tongue position, a wider mouth opening, and a higher concentration of low vowels and back vowels. In bel canto singing, the air column within the resonance cavity is long and thick, and overtones occur frequently, giving the sound a dark, recessed quality. Ethnic singing techniques extensively draw upon and adapt bel canto singing methods in terms of vocal production, breathing, and resonance. Therefore, this paper introduces ethnic singing techniques for comparative analysis.

The vocal samples used in this experiment for both bel canto and ethnic singing techniques were recorded from students who have studied at a music academy for many years, under the guidance of leading professors in China's vocal music education field. The entire experiment was conducted in strict accordance with the procedures of musical acoustics measurement, ensuring the accuracy and reliability of the sound source samples. All sound source samples used in this experiment were recorded in a standard recording studio. Other recording equipment and methods also strictly adhered to the requirements of musical acoustic measurement, effectively ensuring the standardization of the sound source samples.

III. A. Spectral Analysis of Bel Canto Singing Voice

The vocal spectrum of the high-pitched a vowel in the soprano range of a bel canto singer is shown in Figure 3. It can be seen that in bel canto singing, the female voice forms two very clear resonance peaks in the frequency ranges of 2100–2400 Hz and 2900–3200 Hz. It is worth noting that the resonance peak formed in the 2900–3200 Hz frequency range not only has width but also height, with a relatively dense distribution. Additionally, a very clear resonance peak is formed in the 3800–4200 Hz frequency range.

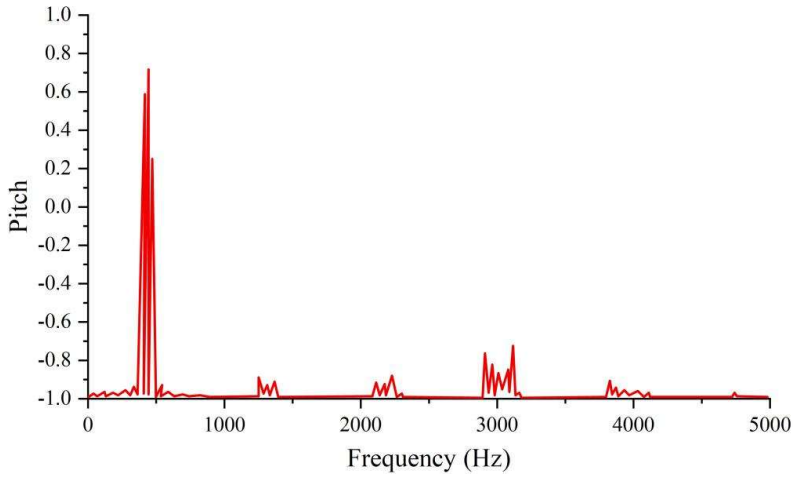

Figure 3: Spectrogram of the vowel sound in the soprano's high-pitched voice range

III. B. Analysis of vocal acoustic parameters for different singing styles

III. B. 1) Analysis of sustained vowel singing

The comparison results of fundamental frequency and velocity quotient between female bel canto singing (code X1) and ethnic singing (code X2) are shown in Figure 4. Overall, as the fundamental frequency increases, the velocity quotient of both singing styles decreases, with the fundamental frequency curve and velocity quotient curve intersecting at a critical point in the low-mid range. Specifically, the velocity ratio for bel canto singing decreases from 363.07% to 134.81%, while the velocity ratio for ethnic singing decreases from 267.65% to 144.91%. The velocity ratio for bel canto singing is significantly higher than that for ethnic singing between the G and C1 notes in the lower octave, but the difference becomes negligible after C1 in the first octave. The opening ratio of both singing styles shows a trend of first increasing and then decreasing. The opening ratio of the bel canto singing style increases from 57.55% to 70.25% and then fluctuates downward to 59.16%. The opening ratio of the ethnic singing style increases from 54.66% to around 62.13% and then fluctuates downward to 56.83%. The opening ratio of the bel canto singing style is slightly larger than that of the ethnic singing style overall.

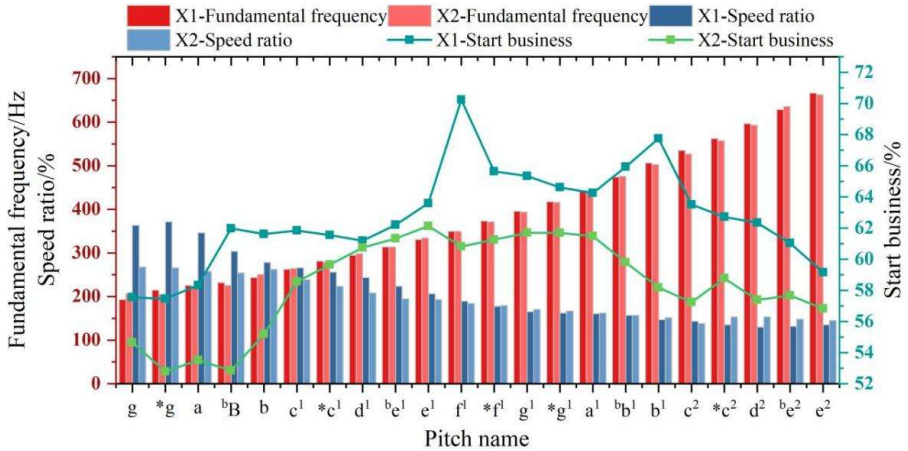

Figure 4: Comparison of Female Opera Singing Style and Ethnic Singing Style

The fundamental frequency, speed ratio, and opening ratio of male bel canto singing and ethnic singing are shown in Figure 5. Overall, as the fundamental frequency increases, the speed ratio of bel canto singing shows a clear

trend of first increasing and then decreasing at the transition point from the low register to the middle register and then to the high register, rising from 334.16% to 403.65% and then decreasing to 216.37%. The speed ratio of ethnic singing also increases from 233.60% to 242.31% and then decreases to 179.53%, but the change is not as pronounced as in bel canto singing. The speed ratio of bel canto singing is generally higher than that of ethnic singing. The overall trend of the opening ratio for both singing styles also shows an initial increase followed by a decrease, but this is not as pronounced as in female voices. The opening ratio of bel canto singing is generally lower than that of ethnic singing.

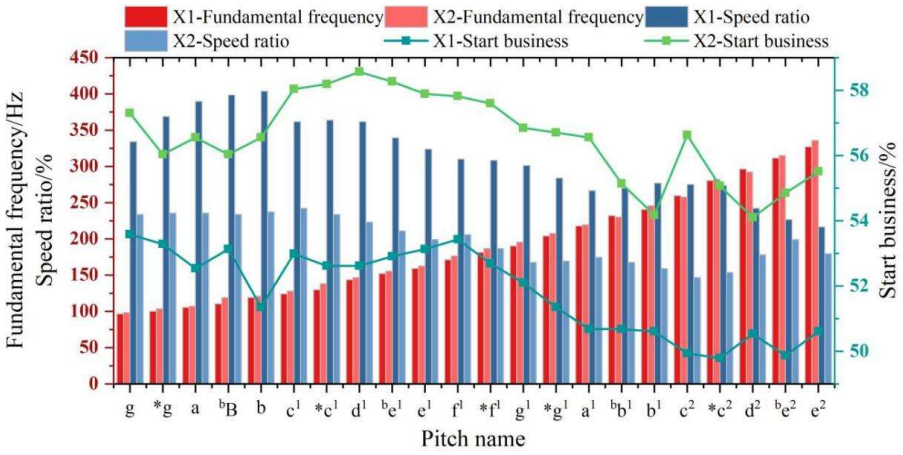

Figure 5: Comparison of Male Opera Singing Style and Ethnic Singing Style

Under normal speaking conditions, as the fundamental frequency increases, the speed quotient generally follows a trend of first increasing and then decreasing, while the opening quotient often exhibits a turning point, as energy changes occur at the transition points between vocal registers. The data in this paper show significant differences in performance between female and male voices when singing sustained vowels. From the perspective of the speed quotient, the speed quotient of female voices generally decreases as the fundamental frequency increases, while the speed quotient of male voices follows a similar trend to normal speech, first increasing and then decreasing. From the perspective of the opening quotient, the speed quotient of female voices first increases and then decreases, while the speed quotient of male voices shows a less pronounced trend of increase and decrease.

III. B. 2) Analysis of vocal ranges in song performance

An analysis of the vocal range characteristics of male and female voices when singing the introduction and main sections of a particular song. The results of the vocal range comparison between the same gender but different singing styles are shown in Figure 6. It can be seen that, regardless of whether it is male or female voices, the overlapping parts of the vocal ranges for both singing styles are relatively large. Compared to the ethnic singing style, the vocal range of the bel canto singing style is slightly larger, specifically manifested in the amplitude of the bel canto female singing style being slightly shifted upward and the fundamental frequency being slightly shifted to the right, meaning that the energy used during singing is slightly greater than that of the ethnic singing style. For male bel canto singing, the fundamental frequency range shifts slightly to the right, but the overall amplitude range shifts slightly downward. This indicates that for female voices, changes in fundamental frequency and amplitude are proportional, with higher fundamental frequencies corresponding to increased amplitude. However, for male voices, as the fundamental frequency increases, the amplitude becomes more concentrated in the middle range.

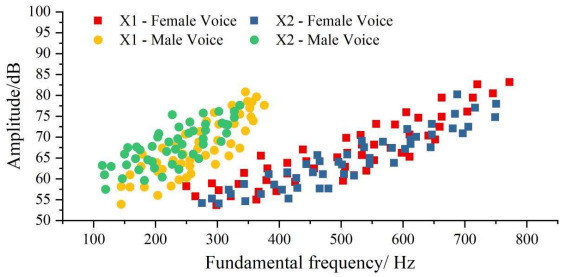

Figure 6: Comparison Results of Vocal Regions

In summary, overall, bel canto singing and ethnic singing exhibit similarities in acoustic parameters such as fundamental frequency, speed ratio, and opening ratio, resulting in a subjective perception of similarity in tone color between the two singing styles.

III. C. Analysis Results of Resonance States in Bel Canto Singing

After recording, the voice teacher screened and analyzed the students' vocal samples one by one, retaining those with errors of “lack of head resonance” and “lack of oral resonance” as the head resonance error group and oral resonance error group in high-pitched vocalization. Using Praat software, the formant peaks of the collected bel canto signals were extracted. The first and second formant peak parameters and standard deviations for bel canto vowel vocalization are shown in Tables 1 and 2, respectively. It can be observed that in the vowels /a/, /u/, and /o/, the F1 and F2 values of the incorrect vocalizations are smaller than those of the standard vocalizations. In the production of these three vowels, the highest point of the tongue position is typically located in the middle-rear portion of the oral cavity, resulting in a smaller mouth opening when the head cavity and oral cavity are not in resonance.

In the vowels /i/ and /e/, the F1 values of the oral group are greater than those of the head cavity group, while the standard pronunciation F1 values fall between the oral and head cavity groups, indicating that the tongue position is higher when the head cavity is not in a resonant state. Conversely, in the F2 values, the oral group is smaller than the head cavity group, but the standard pronunciation F2 values still fall between the two groups. In the formant parameters of standard pronunciation, the F2 value for the vowel /i/ is the highest, primarily because /i/ is a front vowel, which is primarily produced through nasal cavity vibration. The F2 values for the vowels /a/ and /o/ are lower, indicating that these are back vowels, with sound primarily produced through oral cavity vibration. At the same pitch, back vowels are pronounced more clearly than front vowels when singing.

Table 1: Parameters of the First Resonance Peak of Vowels

	Standard pronunciation		No resonance in the head cavity		No resonance in the mouth	
	Average	SD	Average	SD	Average	SD
/a/	818.43	26.38	723.53	80.31	666.24	50.36
/i/	638.52	39.57	502.63	50.22	641.46	48.22
/u/	684.37	18.46	488.12	43.26	563.22	53.61
/e/	617.22	34.33	736.37	46.18	762.18	79.14
/o/	684.55	60.79	510.42	81.37	543.64	75.06

Table 2: Parameters of the Second Resonance Peak of Vowels

	Standard pronunciation		No resonance in the head cavity		No resonance in the mouth	
	Average	SD	Average	SD	Average	SD
/a/	1403.67	53.64	1301.34	98.37	148.35	70.18
/i/	2106.35	70.13	2302.17	218.53	1906.26	235.16
/u/	1594.27	49.37	1104.25	98.94	989.77	109.04
/e/	1936.53	17.44	2019.48	54.27	1805.24	133.53
/o/	1204.86	52.66	933.22	117.36	1002.44	112.44

As shown in the table, the standard deviation of the standard phonation parameters is generally smaller than that of the two error groups, and the formant parameters of the two error phonation groups exhibit significant similarity, poor stability, and substantial interference in phonation classification.

The acoustic vowel distributions of the three groups of bel canto vowel phonations are shown in Figure 7. In the two error phonation groups, the vowel space for phonations with oral issues is smaller, and the distinction between vowel formants is also smaller. The acoustic vowel area of standard phonation is the smallest. Since the F1 and F2 values of the vowel /u/ are both greater than those of the erroneous phonation, the area image does not form an “inverted triangle.” Additionally, the resonant peaks of each vowel in standard phonation are clearly distinguishable, with a smaller standard deviation.

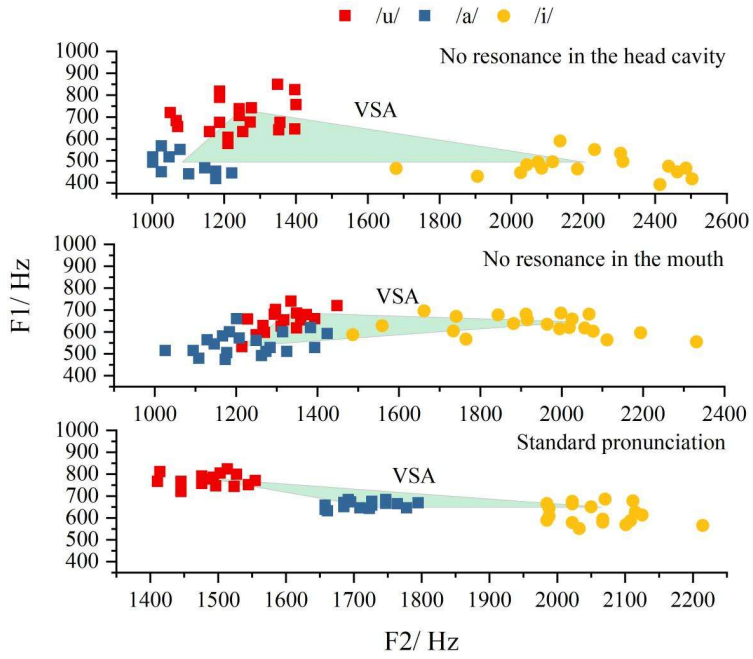

Figure 7: Distribution of Acoustic Vowels

IV. Conclusion

This study systematically investigated the regulatory mechanisms of resonant cavities in bel canto singing using acoustic analysis methods. The main conclusions are as follows.

(1) Acoustic differences between singing styles

The velocity ratio of female bel canto singing decreased from 363.07% to 134.81%, while that of ethnic singing decreased from 267.65% to 144.91%. The open ratio of both singing styles showed a trend of first increasing and then decreasing. The open ratio of bel canto singing increased from 57.55% to 70.25% and then fluctuated downward to 59.16%. The open ratio of ethnic singing increased from 54.66% to around 62.13% and then fluctuated downward to 56.83%. Overall, the open ratio of bel canto singing was slightly greater than that of ethnic singing. The speed ratio of male bel canto singing is generally higher than that of ethnic singing, while the opening ratio is generally lower than that of ethnic singing. In song performance, the vocal range of bel canto singing is slightly wider than that of ethnic singing.

(2) Resonance peak stability and resonance cavity coordination

In the vowels /a/, /u/, and /o/, the F1 and F2 values of incorrect pronunciation are lower than those of standard pronunciation. In the vowels /i/ and /e/, the F1 values of oral group pronunciation are greater than those of the head cavity group, while the F1 values of standard pronunciation fall between the oral group and the head cavity group. Conversely, in the F2 values, the oral group is smaller than the head cavity group, but the F2 values of standard pronunciation still fall between the two groups.

References

[1] Bao, W. (2023). The Inheritance and Development of Italian Bel Canto in Modern Vocal Music Education. Frontiers in Art Research, 5(18).
[2] Yue, C., & Maneewattana, C. (2025). Music Education: The Chinese Bel Canto Theory of Peking University's Opera Research Institute: A Case Study of the Opera Jiang Jie. Journal of Buddhist Education and Research (JBER), 11(1), 326-337.
[3] Hoch, M. (2024). Reframing Bel Canto in the Twenty-First Century: Dovetailing Tradition with Science-Informed Pedagogy. Journal of Singing, 80(4), 417-427.
[4] Mayr, A. (2017). Investigating the voce faringea: Physiological and acoustic characteristics of the bel canto tenor's forgotten singing practice. Journal of Voice, 31(2), 255-e13.
[5] Xie, X. (2021). On the Similarities and Differences Between Bel Canto and Popular Singing. Frontiers, 2(1).
[6] Băjea, M. E. (2021). Vocal Ornamentation in the 18th century-Tosi and Mancini, first Theorists of Bel Canto. Bulletin Of The Transilvania University Of Braşov, Series VIII: Performing Arts, 14(Suppl), 21-32.
[7] Bradford, Z. (2019). Vocal resonance: Optimising source-filter interactions in voice training. Fusion Journal, (15), 47-70.
[8] Plüss, D., & Haspelmath-Finatti, D. (2022). Singing/Embodiment/Resonance. International handbook of practical theology, De Gruyter, 533-546.
[9] Aura, M., Geneid, A., Bjørkøy, K., Rantanen, M., & Laukkanen, A. M. (2022). A nasoendoscopic study of "head resonance" and "imposto" in classical singing. Journal of Voice, 36(1), 83-90.

[10] Meerschman, I., Van Lierde, K., Peeters, K., Meersman, E., Claeys, S., & D'haeseleer, E. (2017). Short-term effect of two semi-occluded vocal tract training programs on the vocal quality of future occupational voice users:“Resonant Voice Training Using Nasal Consonants” Versus “Straw Phonation”. Journal of Speech, Language, and Hearing Research, 60(9), 2519-2536.
[11] Goldstein, J. (2018). Resonance–A key concept in the philosophy of Charles Taylor. Philosophy & Social Criticism, 44(7), 781-783.
[12] Sarkar, M., & Madabhavi, I. (2025). Vocal resonance: a narrative review. Monaldi Archives for Chest Disease.
[13] Wolfe, J., Garnier, M., Bernardoni, N. H., & Smith, J. (2020). The mechanics and acoustics of the singing voice: Registers, resonances and the source–filter interaction. In The Routledge Companion to Interdisciplinary Studies in Singing, Volume I: Development (pp. 64-78). Routledge.
[14] Rakerd, B., Hunter, E. J., & LaPine, P. (2019). Resonance effects and the vocalization of speech. Perspectives of the ASHA special interest groups, 4(6), 1637-1643.
[15] Zhang, P. (2021, November). The Effect of Cavity Resonance on the Timbre of Vowels. In 2021 3rd International Conference on Literature, Art and Human Development (ICLAHD 2021) (pp. 533-536). Atlantis Press.
[16] Learning, L. (2021). Sound Interference and Resonance: Standing Waves in Air Columns. Fundamentals of Heat, Light & Sound.
[17] Liu, J., Wang, T., & Chen, M. (2021). Analysis of sound absorption characteristics of acoustic ducts with periodic additional multi-local resonant cavities. Symmetry, 13(12), 2233.
[18] Wade, L., Hanna, N., Smith, J., & Wolfe, J. (2017). The role of vocal tract and subglottal resonances in producing vocal instabilities. The Journal of the Acoustical Society of America, 141(3), 1546-1559.
[19] Zhang, X., He, N., Cheng, L., Yu, X., Zhang, L., & Hu, F. (2024). Sound absorption in sonic black holes: Wave retarding effect with broadband cavity resonance. Applied Acoustics, 221, 110007.
[20] Kittimathaveenan, K., Thumwarin, P., Pakalong, N., & Nundrakwang, S. (2018, July). Resonance in Vocal Techniques Analysis. In 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST) (pp. 1-4). IEEE.
[21] Raveendran, R., & Krishna, Y. (2023). Auditory-perceptual judgment of vocal resonance in Carnatic singers by different groups of listeners. Journal of Voice.
[22] Cai, H. (2019, October). Acoustic analysis of resonance characteristics of head voice and chest voice. In 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-6). IEEE.
[23] Köberlein, M., Birkholz, P., Burdumy, M., Richter, B., Burk, F., Traser, L., & Echternach, M. (2021). Investigation of resonance strategies of high pitch singing sopranos using dynamic three-dimensional magnetic resonance imaging. The Journal of the Acoustical Society of America, 150(6), 4191-4202.
[24] Nair, A. (2021). The tongue as a gateway to voice, resonance, style, and intelligibility (Vol. 1). Plural Publishing.
[25] Vos, R. R., Daffern, H., & Howard, D. M. (2017). Resonance tuning in three girl choristers. Journal of Voice, 31(1), 122-e1.
[26] Vurma, A. (2022). Amplitude effects of vocal tract resonance adjustments when singing louder. Journal of Voice, 36(2), 292-e11.
[27] do Amaral Ambros, G., & e Silva, M. A. D. A. (2025). Resonance Strategies in the Upper Range of Western Operatic Tenors. Journal of Voice.
[28] Jeanneteau, M., Hanna, N., Almeida, A., Smith, J., & Wolfe, J. (2022). Using visual feedback to tune the second vocal tract resonance for singing in the high soprano range. Logopedics Phoniatrics Vocology, 47(1), 25-34.
[29] Yamasaki, R., Murano, E. Z., Gebrim, E., Hachiya, A., Montagnoli, A., Behlau, M., & Tsuji, D. (2017). Vocal tract adjustments of dysphonic and non-dysphonic women pre-and post-flexible resonance tube in water exercise: a quantitative MRI study. Journal of Voice, 31(4), 442-454.