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

A study on the dynamic adjustment mechanism of resonance cavity in American vocal singing based on acoustic analysis

Guili Yao[1,*]

[1] School of Arts, Shandong University (Weihai), Weihai, Shandong, 264200, China

SUMMARY: *A study was conducted to capture the vocal signals of eight American singers as subjects under multiple tasks (vowel scales, self-defined pitch exercises, given pitches, song singing and natural speech). And seven categories of core acoustic parameters were extracted. The parameters include the resonance peaks that reveal the shape of the cavity, the fundamental frequency that determines pitch, the range that reflects the capacity of the vocal folds, the fundamental frequency perturbation and the resonance peak perturbation to assess stability, and the average energy that represents the intensity of the sound. For the vowel /a/ in male voices, the first resonance peak peak value shifted leftward from the standard 717 Hz to 589 Hz in the absence of oral resonance, whereas the third resonance peak F3 of /u/ in female voices fell from the standard vocalization of 3088 Hz to 1990 Hz in the absence of head resonance. In the absence of head resonance, the area of the vowel space of the female voice, VSA, was expanded to 177,762 Hz², which is 48 times of the standard value, and the resonance concentration rate, FSA, is 48 times the standard value. Times the standard value, while the resonance peak concentration ratio FCR decreased from 1.682 to 1.405.Once the support of the critical resonance cavity is lost, the vocal tract modulation becomes misaligned, resulting in a significant loss of articulatory clarity.*

KEYWORDS: *acoustic analysis; American singing; resonant cavity; resonance peaks; acoustic parameters*

1 Introduction

American vocal singing is different from other singing styles in terms of resonance, which is to mobilize all the resonance cavities that can be used in singing to obtain overall resonance [1]. That is to say, we train the method of adjusting the resonance cavities, appropriately expanding the resonance organs that can be adjusted in shape - pharyngeal, oral and laryngeal cavities, and using the resonance organs that cannot be adjusted in shape - head and chest cavities, so as to increase the volume of sound and beautify the timbre through the adjustment and use of the resonance cavities.

The resonating organs of the human voice are vital to the sound production of the voice. The oropharynx, which is the resonating organ of the human voice, can change the internal shape and spatial volume of the oropharynx through the regulation of the physiological movement of the muscles, so as to produce various changes in thickness, depth, roundness, flatness, size, etc., which can change the sound reflex performance of the oropharynx and enhance the volume of the voice [2-4]. It is also possible to adjust the resonance function and sound position of different cavities through the resonance connection of different resonance

*lsxieping@163.com
https://doi.org/10.65102/is2026518
Published: 30 April, 2026

cavities, so as to beautify the timbre, change the voice farsightedness and voice penetration [5-7]. Modern scientific research has proved that the so-called resonance organs of the human body mentioned in the traditional vocal theory, including the thoracic cavity, cephalic cavity, nasal cavity, sinus, etc., are fixed and unadjustable [8]. They do not play a role in singing vocalization and resonance production. Therefore, the regulation of singing resonance starts from the singing organs that affect vocalization and resonance. Mouth, tongue, soft palate, pharynx and larynx, i.e., the organs and tissues in the vocal tract are variable and can be governed and regulated by human consciousness and ability, and they are the organs that have an important influence on the quality of vocalization and resonance and its propagation and play a key role in it [9-13]. The method of resonance regulation in American voice singing focuses on the effective operation of the relevant parts. Dynamic regulation of the resonance cavity in American voice singing helps to improve the quality of the voice, vocal expression, and the unity of the vocal range. Therefore, it is necessary to sort out the mechanism of dynamic regulation of the resonance cavity in American voice singing and supplement the knowledge of integration of physiological and acoustic dimensions of American voice singing in order to better utilize the cavity regulation methods.

In order to explore the dynamic rules between human resonance cavities such as oral cavity and pharyngeal cavity and acoustic output behind different vocal range transitions. The study takes American voice singing as an entry point and utilizes acoustic analysis to explore its dynamic regulation mechanism. The first step is to capture the different sound states of American vocal singing under controlled conditions, and to extract the physical indicators that can reflect the resonance regulation from these sound signals. In this study, eight professional male and female American vocalists were invited to capture their individualized vocal signals based on basic vowel scale exercises, self-defined pitches, given pitches and flat speech sounds. Based on acoustic theory, the most relevant core features of resonance modulation, such as resonance peaks (the first and third resonance peaks), fundamental frequency (which determines pitch), and range, were selected for analysis. Combining fundamental frequency perturbation and resonance peak perturbation, subtle fluctuations in frequency and resonance of the sound are quantified. The resonance efficiency is also reflected based on the average energy of the sound. Among them, the first and third resonance peaks were detected by AR model peak detection method, and the fundamental frequency was extracted by autocorrelation combined with wavelet method.

2 Acoustic Experimental Design and Multi-Parameter Acoustic Analysis for American Voice Singing

In order to systematically investigate the dynamic regulation mechanism of the resonance cavity in American voice singing, this study firstly designed relevant acoustic and physiological experiments, aiming at obtaining high-quality singing and speech data under repeatable and comparable conditions.

2.1 Acoustic and Physiological Approaches to the Study of American Voice Singing

2.1.1 Subjects

Since tenor and soprano cover the widest range of registers, it is easy to study the changes of singing voice in different vocal regions, therefore, eight articulators were selected for this paper, and the soprano voice was chosen as American tenor and American soprano, with four articulators in each category. (average age 22.76±1.82 years), and each of them had been paid the corresponding labor fee and signed a data use agreement. The recording sessions chosen were based on each subject's self-reported time of optimal voice condition.

2.1.2 Experimental tasks

(1) Vowels

Vowels are the most musical of speech sounds, with clear pitch and rich harmonic structure. The main way to distinguish between different vowels within a language is phonological quality, and the fact that there is always a change in F0 for vowels produced individually is one of the reasons why vowels have a speech sound quality. the pattern of F1 and F2 is the most important cue for recognizing vowels. When moved far enough, they can even change their linguistic identity. At the same time, the choice and tuning of different vowels are also crucial in distinguishing singing styles.Therefore, vowels are both an important object of phonological research and the basis of singing vocalization, and one of the bridges between language and music.The study of vowels in singing voice and speech voice is very necessary.

1) Routine vocalization exercises

Subjects performed the regular pentatonic ascending and descending vocalizations of the vowels $/a/$, $/i/$, and $/u/$, respectively. The details are as follows: first, the computer played the pre-recorded “do-mo-sol (relative pitch)” chord as a piano cue, and then the subjects performed the “do-re-mi-fa-sol-fa-mi-re-do (relative pitch)” vocalization exercise. The vocalization of “do-re-mi-fa-sol-fa-mi-re-do (relative pitch)” was then practiced, one semitone at a time, starting at the lower limit and ending at the upper limit of the subject's range, and once for each vowel corresponding to the pitch of the vowel. 70 vocalizations were performed by each of the 8 subjects, and the task consisted of 560 samples.

2) Self-pitched exercises

Subjects performed self-defined pitch exercises with the vowels $/a/$, $/i/$, and $/u/$, respectively, as follows: subjects vocalized according to self-perceived intervals of the range with the lower and upper limits of the range, and each vowel corresponding to the pitch was repeated two times, respectively. For example, the American tenor sang 20 sets of vowels $/a/$ in the range 224 Hz-983 Hz, 16 sets of vowels $/i/$ in the range 134 Hz-893 Hz, and 18 sets of vowels $/u/$ in the range 202 Hz-955 Hz. The samples used for the analysis of the self-pitching exercise were 432 ((20+16+18)*8).

3) Given Pitch Exercise

The subjects performed the given pitch exercises (given by the computer software) with the vowels $/a/$, $/i/$, and $/u/$ as follows: due to the difference in innate physiological conditions, female voices are one octave higher than male voices, therefore, the given pitch exercises were needed by gender. The given pitches for female voices were 387Hz, 515Hz, 623Hz and 689Hz, and the given pitches for male voices were 189Hz, 246Hz, 306Hz and 379Hz, and the pitches corresponding to each vowel were repeated twice, and a total of 192 samples (3*8*8) were collected and analyzed.

4) Speech sounds

Subjects pronounced the vowels $/a/$, $/i/$, and $/u/$ continuously for 3 to 5 seconds in a loaded Mandarin yinping tone under normal speech condition, and each vowel was repeated twice. Speech sounds were collected and analyzed in 54 samples (9 per pronouncer)

(2) Songs

Songs were mainly used for the study of singer resonance peaks. In terms of the song selection criteria, firstly, it was ensured that the songs were familiar and well known to the subjects, and there could be differences in the repertoire between the subjects. American vocalists chose foreign classic songs to sing. Eight subjects chose a total of 16 songs. Each song was recorded twice with the key number fixed. The lyrics of the songs were read twice in a normal speaking state.

2.1.3 Recording environment

The recordings were made in a soundproof room for voice laboratory recordings with a background noise of 6.5 dB. The recording software was LabChart 8.2.14, and four-channel recording was used. one-channel speech signals were recorded by using a lavalier condenser microphone (SONYECM-44B), and the distance from each speaker's lips to the microphone was about 15-22 cm. two-channel electronic gates were collected by an electronic gatesmeter (TG2-PCX2) to capture electronic gatesmeter signals (EGG), three-channel for chest respiratory signal, four channels for abdominal respiratory signal, of which the sampling frequency of one and two channels is 40.0kHz, three and four channels are 1.0kHz, and the four signals enter into the laptop computer through the Powerlab Multi-Conductor Physiological Recorder. In this study, the malefactor extracted one channel of voice signal and two channels of electronic sound gate meter signal. The sound pressure level (SPL) was calibrated by a pure tone at 1000 Hz. This was done by placing an AWA5715 sound level meter at the lips of the pronouncer, ensuring that the distance between the microphone and the sound level meter and the distance of the microphone from the lips were the same, and finally calibrating based on the difference between the data measured in the field and the data measured in the Prrat.

2.2 Parameter rotation and extraction based on acoustic theory

This section provides an overview of the basic features of the acoustics of American singing, on the basis of which a system of acoustic parameters closely related to the dynamic regulation of the resonating chamber is filtered.

2.2.1 Selection of Acoustic Parameters for American Voice Singing

(1) Resonance Peak

Resonance Peak is a doctrine that studies the resonance and sound quality of the voice. Sound quality, volume and timbre are three important components of the voice. The low resonance peaks or frequency distribution of the voice provide depth or resonance, and the high resonance peaks provide brilliance. All of these qualities come together to form the voice's tonal quality. In acoustic terms, tone quality is the spectral properties of vowels as reflected in the spectrograph. The quality of the sound determines the timbre and volume characteristics of the voice. Therefore, the study of singing, should first focus on the study of the quality of sound, without good singing sound quality, volume and timbre can not be discussed. Without good singing voice quality, there is no way to talk about volume and timbre. It is also wrong to equate voice quality with timbre.

Generally speaking, the lower of the two peaks that is the first resonance peak and the second resonance peak basically stipulates the sound of the vowel timbre, while the higher of the third resonance peak, the fourth resonance peak and the fifth resonance peak affects the sound of the personal characteristics and musical timbre. The value of the intrinsic frequency of the vocal tract is determined by the morphology of the vocal tract, and changing the morphology of the vocal tract produces a different sound, so each vocal tract morphology is characterized by a set of resonance peaks.

Research results show that the first and third resonance peaks are important objective reference data to measure the vocal skill level of singers. In this article, the first and third resonance peaks are chosen as objective evaluation parameters.

(2) Fundamental frequency

Fundamental frequency is the basic frequency of vocal folds vibration when producing turbid voice; in addition to the basic characteristics of the vocal folds (length, quality, tension, etc.), it is also regulated by the cricoarytenoid muscle, arytenoid muscle, and the downward pressure of the vocal folds; it is one of the most important parameters of the singing voice.

(3) Range

Tone range refers to the range of human vocal ability from low to high in pitch. The range can be divided into five types: physiological range, vocal range, natural range, singing range, and total range of the human voice.

Singing range is determined by two conditions, one is whether the vocal folds can emit vibration frequency broad “sound source”; the second is whether the resonance organ can be adjusted to the same as the vocal folds vibration frequency of the same or similar to the state of resonance. Vocal cords are two pieces of flexible muscle, there are thick and thin, loose and tight changes. According to the principle of acoustics, "long, large, thick, thick and tension of the more relaxed object vibration is slow, low frequency. Conversely, shorter, smaller, thinner, thinner and more tense vocal bodies vibrate quickly and with a high frequency." The vocal cords are directed by mental intention to make changes in length, thickness, thinness, and looseness, realizing changes in the height of the voice.

Singing range reflects an aspect of an actor's singing ability and is an important indicator of American voice research.

(4) Base frequency perturbation

Fundamental frequency perturbation refers to the rate of change of the fundamental frequency of sound waves between neighboring cycles, which is used to measure the amount of difference between a specified cycle and the neighboring previous or subsequent cycles, reflecting the difference in frequency between the periodicity of vocal fold vibration.

(5) Resonance peak perturbation

The first and third resonance peak perturbations are measures of the rate of change of the first and third resonance peaks between adjacent cycles, respectively.

(6) Average Energy

The average energy indicates the relative size of the vocal signal in the same environment.

2.2.2 Extraction methods

(1) Resonance peak extraction methods

The methods for extracting the resonance peaks include the band-pass filter bank method, the cepstrum method, the LPC valley detection method, and the AR model peak detection method.

The first and third resonance peaks are extracted by the AR model peak detection method. The extraction process is shown in Fig. 1.

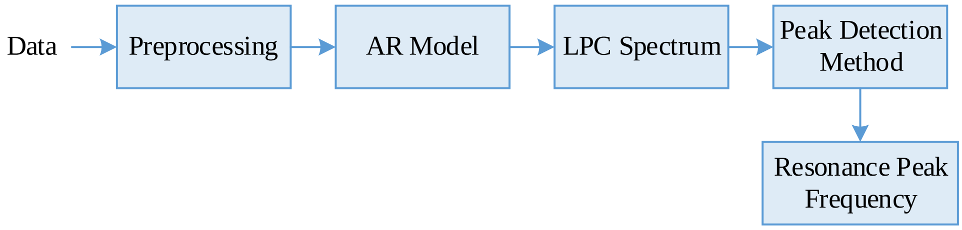


Figure 1: AR model peak detection method

(2) Base frequency extraction methods

The methods of extracting fundamental frequency include autocorrelation function method, average amplitude difference function method, wavelet transform method, and the method combining autocorrelation and wavelet.

The experiment chooses the method combining autocorrelation and wavelet to extract the fundamental frequency. The steps of fundamental frequency extraction are:

1) Song signal reading;

2) Split frame;

3) Db4 wavelet decomposes each frame signal in 4 layers respectively;

4) Reconstruction of the decomposed wavelet signal;

5) Autocorrelation calculation;

6) Peak detection. The frequency corresponding to the peak position is the fundamental frequency.

(3) Tone range extraction method

The simplest way to estimate the pitch range is to select the maximum and minimum values of the pitches appearing in the songs and scores, but there are many pitch data in the songs and scores, and it is inevitable to take only two of them by chance, so it is necessary to use the statistical method. The method of estimating the pitch range in this experiment is to take the average and standard deviation of all the pitch D values in the song and songbook.

1) Calculation of pitch D values

Pitch is determined by the vibration frequency of the object (vocal cords). That is, in a certain unit of time, the object vibration number of times, then the pitch; vibration number of times, then the sound is low. Tone and intonation are both manifestations of pitch. Tone, after all, is not a direct reality of the fundamental frequency, but an abstraction cut across from the Hertz-based pitch curve. Therefore, the experiments depict pitch in terms of the D value, which is the logarithmic scale of pitch, defined as:

$$D = 12 * \log_2 \left(F \div F_0 \right) \tag{1}$$

The pitch F in Hertz is the step difference of the reference frequency F_0. The value of D is a dimensionless number whose unit can be taken as “degrees”, and is replaced by D for ease of writing.

Definition of 2 as the base of the logarithm, that D value is proportional to the range; constant 12 means that every difference of an octave, D value minus 12, and the piano scale just in line. In the experiment, 80 Hz was used as the reference frequency to calculate the “absolute” pitch. This F_0 value was chosen to avoid using negative values for pitch, and to keep the male and female pitch values close to each other.

The amplitude of the vocal pitch (pitch) variations reflected directly by the Hertzian values is very different, but the amplitude of the variations reflected by the D values is very similar; and the difference in the D values is independent of the gender. From the above aspects, it can be seen that the D value is a more practical measurement method derived from the experience of former people for the need of pitch measurement.

(2) The average of all pitches in a song or a score $\bar{D}$

Take the average and standard deviation of all the pitch D values in the song and the score:

$$\bar{D} = \frac{1}{N} \sum_{j=1}^{N} D_j \tag{2}$$

3) Standard deviation σ

$$\sigma = \sqrt{E[(D_j - \bar{D})^2]},\ j = 1, 2, \cdots, N \tag{3}$$

where $E[]$ denotes averaging and N is the number of pitch data elements.

4) Domain width S

$$S = 4\sigma \tag{4}$$

(4) Base frequency perturbation extraction method

Mathematical definition of fundamental frequency perturbation:

$$jitter = \frac{1}{N-1}\sum_{i=1}^{N}\left|1/F_{0i} - 1/F_{0(i-1)}\right| \tag{5}$$

where F_{0i} denotes the fundamental frequency of the ith cycle.

(5) Resonance peak perturbation extraction method

Mathematical definition of resonance peak perturbation:

$$\text{First Resonance Peak Perturbation} = \frac{1}{N-1}\sum_{i=1}^{N}\left|1/F_{1i} - 1/F_{1(i-1)}\right| \tag{6}$$

$$\text{Third Resonance Peak Perturbation} = \frac{1}{N-1}\sum_{i=1}^{N}\left|1/F_{3i} - 1/F_{3(i-1)}\right| \tag{7}$$

where F_{1i}, F_{3i} denote the first resonance peak and the third resonance peak of the ith cycle, respectively.

(6) Average energy extraction method

Mathematical definition of short-time average energy:

$$E_n = \sum_{k=-\infty}^{+\infty} x^2(k)w(n-k) \tag{8}$$

3 Experimental data analysis of resonance cavity regulation in American voice singing

Based on the experimental setup and the acoustic parameters on the extraction method, Chapter 3 analyzes the experimental data to explore the dynamic regulation mechanism of the resonance cavity under American singing.

3.1 Spectral characteristics of vowel vocalizations

After the acoustic parameters were systematically collected and extracted, attention was first focused on the most basic singing unit vowels. Its spectrum was analyzed to establish an overall impression of American singing in terms of acoustic performance, and the characteristic differences between male and female vocal parts were explored in depth.

3.1.1 Analysis of tenor and soprano vocalization spectra

For the experiment of American singing subjects, take the vowel $/a/$ as an example, one tenor and one soprano were selected to analyze the vocalization spectrum of the /a/ vowel in their high voice region, and the vocalization spectra of the two are shown in Fig. 2.

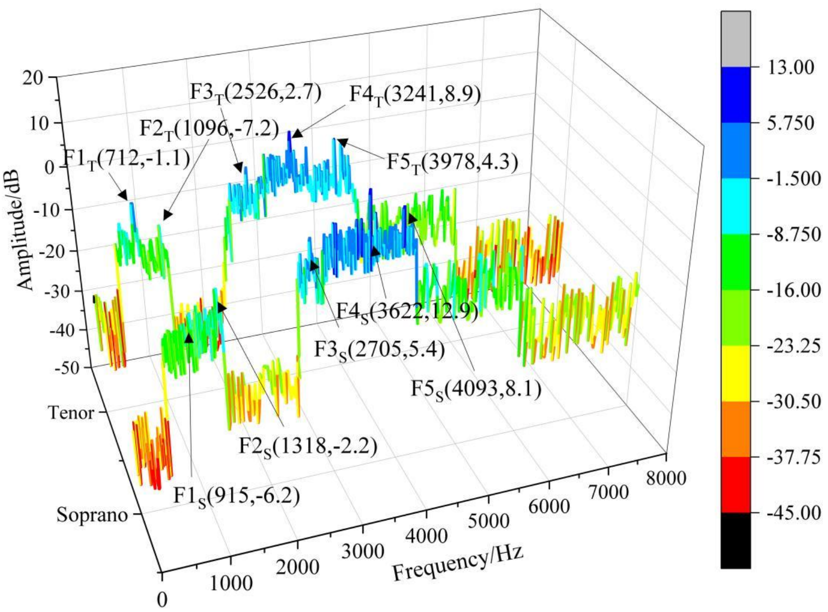


Figure 2: The vocal frequency spectra of the /a/ vowel for tenors and sopranos

In American singing, the focus is on the tenor spectrum (slightly back of the curve), which has a fundamental frequency roughly around 440Hz (standard A4 tone). You can see from the graph that small spikes of energy appear every approximately 440 Hz in the low frequency region from 0-2000 Hz. The tenor resonance peaks F1 and F2 occur at 400-800Hz and 800-1200Hz respectively, with peaks around 712 and 1096Hz. Moving further to the right to the 2500-4200Hz range, where the amplitude is >0, the singer's resonance peak F3 occurs, augmenting a very broad energy wave frequency that peaks at about 3200Hz.

The soprano's fundamental frequency is higher, at about 880Hz (A5 tone), so her spectral curve has a wider harmonic interval and a higher starting point. Her first and second resonance peaks are also higher than the tenor's, with F1 at 915Hz and F2 at 1318Hz, making her vowels sound brighter. At the same time, two very clear resonance peaks are formed in the frequency range of 2500-2800Hz and 3300-3700Hz respectively, with the peaks around 2705Hz and 3622 Hz, and the resonance peaks formed in the frequency range of 3300-3700Hz are wide and high, and densely distributed. In addition to the frequency domain value in the range of 3700-4300 Hz will also form a very clear resonance peak, resonance peak F5 peak at about 4100 Hz.

3.1.2 Spectral envelope analysis

Based on the above obtained spectral analysis of the tenor soprano's vocalizations at the /a/ vowels, a spectral envelope analysis was performed in order to show more clearly the characteristics of their resonance peaks. The spectral envelope is the curve formed by connecting the amplitude vertices of each frequency component in the spectrogram of Fig. 2, which is used to characterize the overall shape and trend of the spectrum.

The spectral envelopes of the tenor and soprano are shown in Figures 3 and 4, respectively.

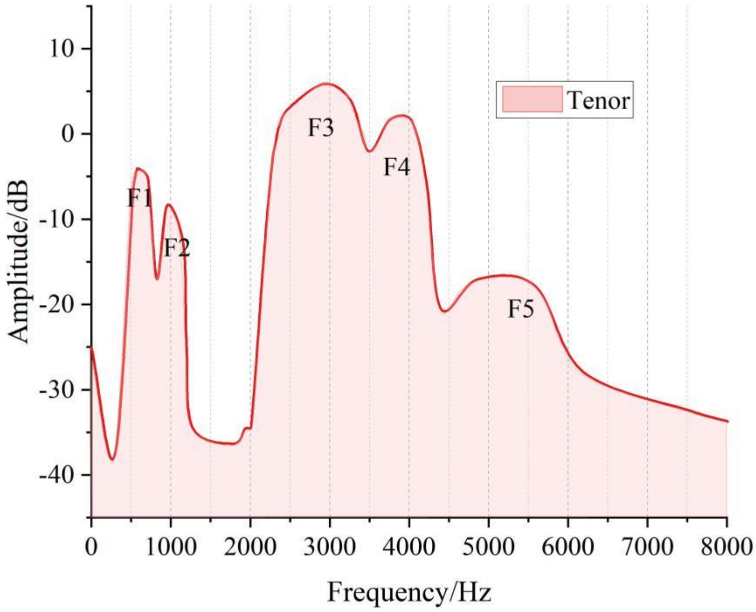


Figure 3: The spectral envelope of the tenor voice

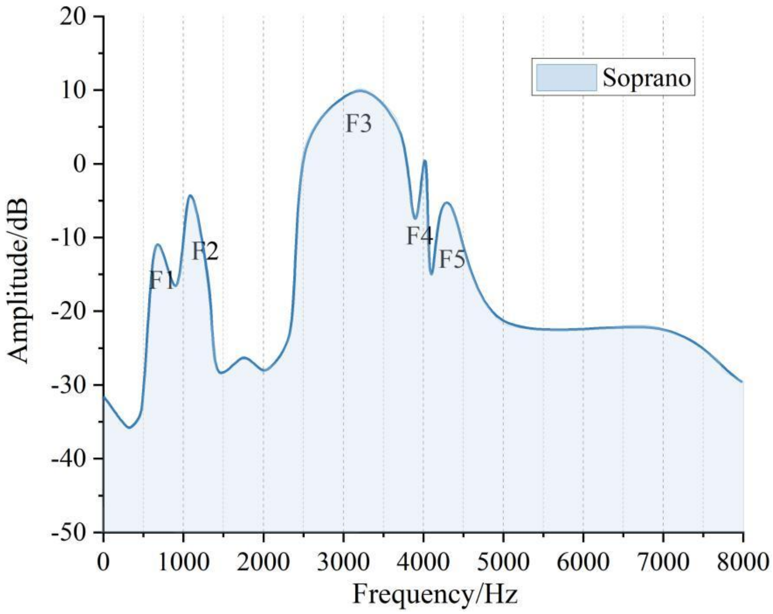


Figure 4: The spectral envelope of the tenor voice

Spectral envelope diagram more clearly show the resonance peaks of male and female voices, focusing on the red curve of the tenor in Figure 3, respectively, the first and second resonance peaks near 700Hz and 1100Hz, followed by a steep rise around 2500Hz, the peak at about 3200Hz, the amplitude of the entire spectrum to reach the apex of the region that is often referred to as the singer's resonance peak in the American vocal singing, is the male singer's energy gathering place. The curve also has a clear but low rebound near 4000Hz (resonance peak F5).

As described in section 3.1.1 Spectral Analysis, the soprano has a higher fundamental frequency (about 880 Hz), and the entire center of gravity of the curve is then shifted up one step from the male voice. Her first resonance peak F1 and second resonance peak F2 peak at 900 Hz and 1300 Hz, respectively, are both positioned higher than the male voice, allowing her to sing the /*a*/ vowels brighter and more forward than the male voice. Between 3300Hz and

3700Hz, the curve rises to a broad and high peak cluster F3, which exceeds the energy of the tenor's peak, with an amplitude of 10dB. The curve continues to rise to independent peaks (resonance peaks F4 and F5) around 3900Hz and 4100Hz, respectively. The energy continues to extend above 6000 Hz, decaying very slowly compared to the male voice.

3.2 Analysis of energy distribution of different tracks for male and female voices

The perspective is now shifted to song singing to analyze the audio singing energy of the male and female voices for different songs in the American voice. For two songs, Mozart's “Avenging Flame” and Puccini's “No One Sleeps Tonight”, the audio singing energy was recorded for both male and female voices in one-second measurements, and their distributions are shown in Figures 5 and 6, respectively.

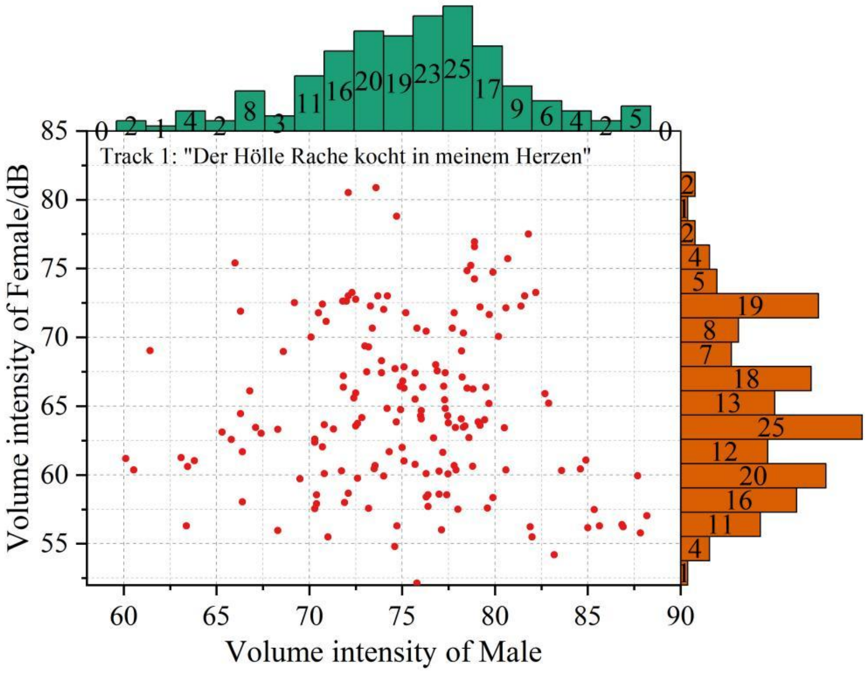


Figure 5: Analysis of the male and female voices in Track 1

Track 1, “The Flame of Vengeance”, was sung by male voices for a total of 2 minutes and 57 seconds, and by female voices for 2 minutes and 48 seconds, with 177 and 168 data points, respectively. By Figure 5 horizontal axis data points can be seen in the male voice singing intensity concentration were between [70,80] dB, up to 88.19dB, as low as 60.11dB, an average of 75.31dB, the repertoire is more impassioned and surging, the male voice sings consistently high pronunciation; while the female voice sings a more dispersed distribution of energy, singing energy in the 55-75dB between the high-low-high-low stepwise distribution, staccato, the lowest 52.14dB, up to 80.88dB, an average of 64.86dB, the overall intensity of the sound is more convergent than the male high voice.

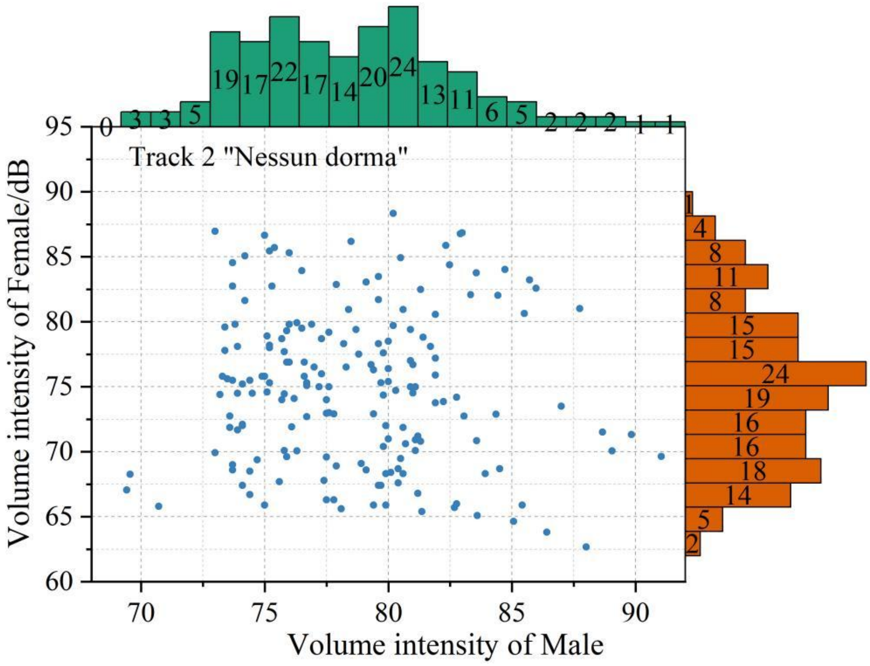


Figure 6: Analysis of the male and female voices in Track 2

Track 2 Puccini's “No One Sleeps Tonight” is sung by men and women for 3min07s and 2min52s respectively, with a total of 187 and 172 data points. Track 2 is more emotionally aggressive than track 1, so the singing energy is more turbulent, higher energy, mainly gathered in the [73,82]dB, the maximum value even went to 91.04dB; the female voice distribution is also more decentralized, with more than 30% of the data points centered in the 65-75dB range. The female voice shapes the emotional delivery of the track through the more frequent use of weak and half voices, with a maximum singing energy of 88.34 dB.

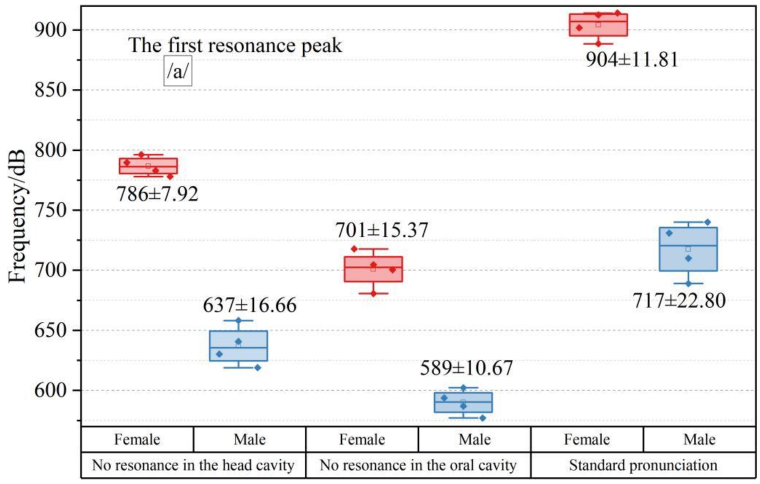
3.3 Resonance peak parameter analysis

Returning to the extracted acoustic parameters, this section provides an in-depth analysis of the first and third resonance peak parameters to understand the mechanisms by which the resonance cavity is regulated in this regard.

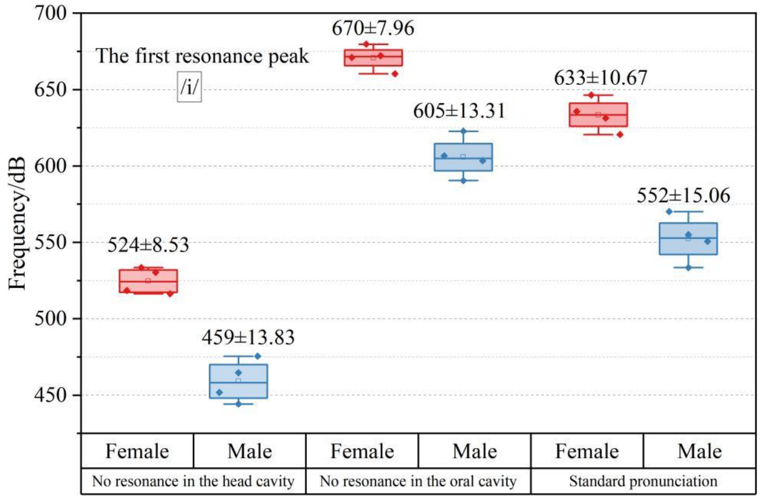
3.3.1 Resonance peak frequency analysis

The resonance peaks were extracted from the acquired American vocal signals of the eight subjects. Figures 7 and 8 show the frequency values of the first and third resonance peaks of /a/, /i/, and /u/ at the occurrence of the first and third resonance peaks under standard vocalization, head cavity unresonance, and oral cavity unresonance, respectively.

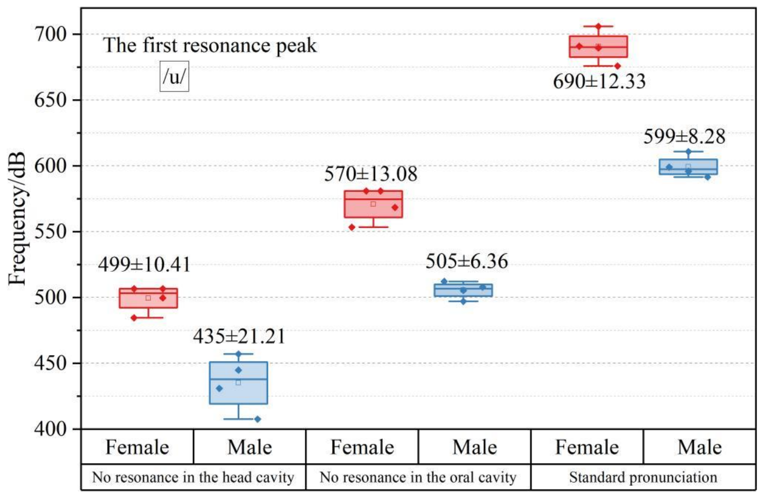
(1) First resonance peak



(a)Vowel /*a*/



(b)Vowel /*i*/



(c)Vowel /*u*/

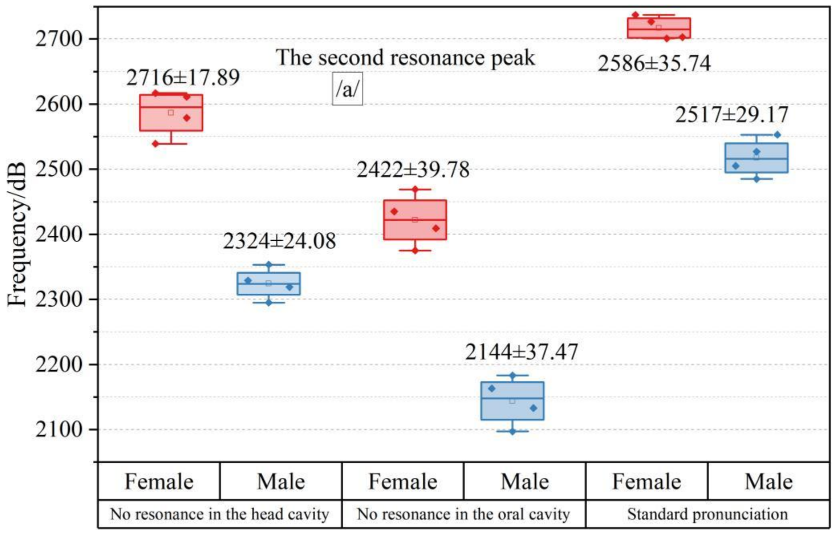
Figure 7: The parameters of the first resonance peak of the trisyllabic vowels

In the three vowels, the frequency change of the first resonance peak F1 conforms to the overall acoustic law that the larger the opening, the higher the F1, while different error resonance states have different effects on the vocal tract morphology. On the /a/ vowels, the average F1 of the eight subjects' male and female voices was 717 Hz and 904 Hz, respectively, during standard vocalizations. Retracing Figure 1, we can see that the first resonance peaks of a particular male and female voice appeared at 712 Hz and 915 Hz, respectively. In the absence of head and oral resonance, the resonance peaks are shifted to the left, falling to 637 Hz and 589 Hz for the male voice and 786 Hz and 701 Hz for the female voice, respectively. In the absence of resonance, the singer unconsciously narrows the mouth opening and elevates the tongue position, which results in a narrowing of the front part of the vocal tract, and thus a reduction in the frequency of the peaks.

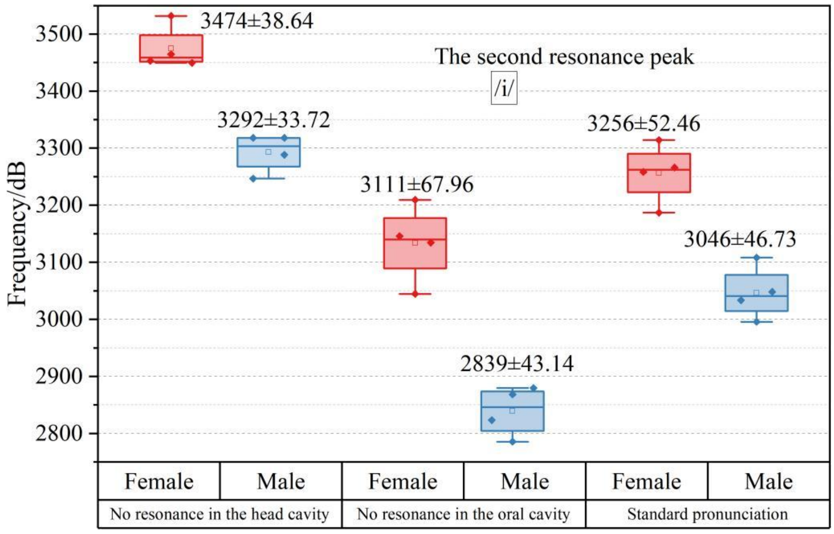
The situation is different for the /i/ vowel, where the first resonance peak frequency in the oral unresonated state is higher than in the standard vocalizations, 605 Hz > 552 Hz for males and 670 Hz > 633 Hz for females, because /i/ is a high, anteriorly positioned vowel, and the tongue may be retracted in the standard vocalizations, which in turn increases the anterior oral volume, resulting in peaks that occur only at higher frequencies. The reason for the low F1 frequency when the head cavity is not resonating is that the tongue position is abnormally elevated and the vocal tract is more constricted when the head cavity support is lost.

On the /u/ vowel, the first resonance peaks peak at 599 ± 8.28 Hz and 690 ± 10.67 Hz for men and women, respectively, during standard vocalizations, whereas when the head cavity is not resonating, the peaks are already present at frequencies of 435 Hz and 499 Hz because the /u/ vowel is highly dependent on the nasal cavity resonating with the head cavity. Once the cephalic resonance support is lost, the vocal tract is extremely tightened and the tongue position is abnormally elevated in order to produce this vowel. In contrast, the oral unresonated state has relatively little effect.

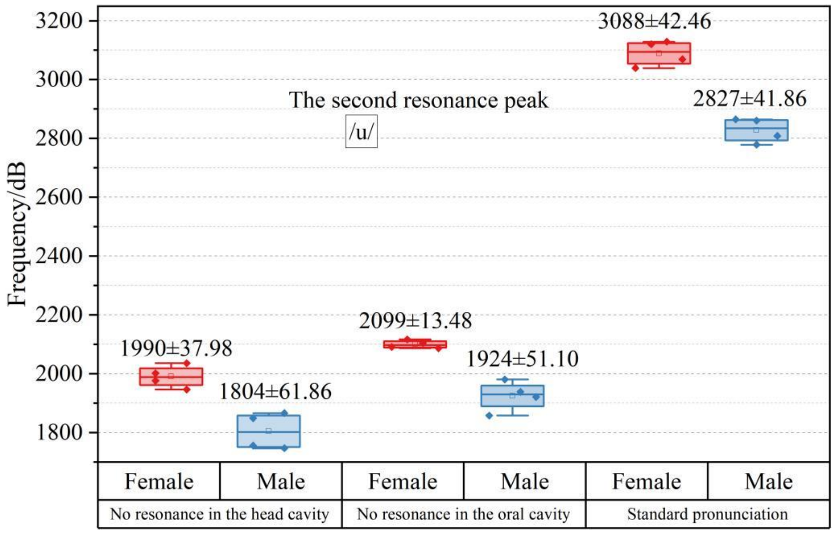
(2) Second resonance peak



(a)Vowel /a/



(b)Vowel /*i*/



(c)Vowel /*u*/

Figure 8: The parameters of the third resonance peak of the trisyllabic vowels

As mentioned earlier in Section 2.2, the first resonance peak (and second resonance peak) specifies the vowel timbre of the voice, while the higher third resonance peaks (fourth and fifth resonance peaks) influence the individual character of the voice and the musical timbre. Changes in the third resonance peak are more sensitive to the condition of the high-frequency resonance channel during vocalization.

For /*a*/ vowels, the peak third resonance peaks occur on average at 2517 Hz and 2716 Hz for male and female voices, respectively, in standard vocalizations, and they also occur earlier in erroneous states, such as 2144 Hz and 2422 Hz when the oral cavity is not in resonance, again because the highest point of the tongue position for /*a*/ vowels is usually at the middle and back of the oral cavity, which leads to the headspace, oral cavity is not in resonance, resulting in a smaller opening when the head and mouth are not in resonance.

For the /*i*/ vowel, the third resonance peaks for both male and female voices were shifted to the right at 3473 Hz and 3292 Hz, respectively, in the absence of cephalic resonance support because the front of the vocal tract is forced to be overly tense and thin in order to maintain pitch and volume, which in turn raises the resonance frequency. This also explains why incorrect shouting can result in a shrill, harsh sound.

The frequency at which the /u/ vowel reaches the third resonance peak during standard vocalization is itself high (2827 Hz for males, 3088 Hz for females), suggesting that high-frequency resonance is required to vocalize this vowel. However, when a cephalic unresonance error occurs, the frequency shifts rapidly to the left, with peaks occurring between 1804 Hz and 1990 Hz, respectively. Frequency shifts in excess of 1000 Hz indicate that the absence of headspace resonance has a dramatic effect on the high-frequency overtone structure required for almost /u/ vowels, and the sound inevitably becomes dull and hollow.

3.3.2 Acoustic spatial parameter analysis

Further, for the vowel space VSA enclosed by the three vowels /a/, /i/, and /u/ in the three states was analyzed, and the acoustic vowel maps for the standard vocalizations of male and female voices, the oral unresonated group, and the cephalic unresonated group are shown in Figures 9, 10, and 11, respectively. (The horizontal axis is the first resonance peak frequency F1, and the vertical axis is the third resonance peak frequency F3.)

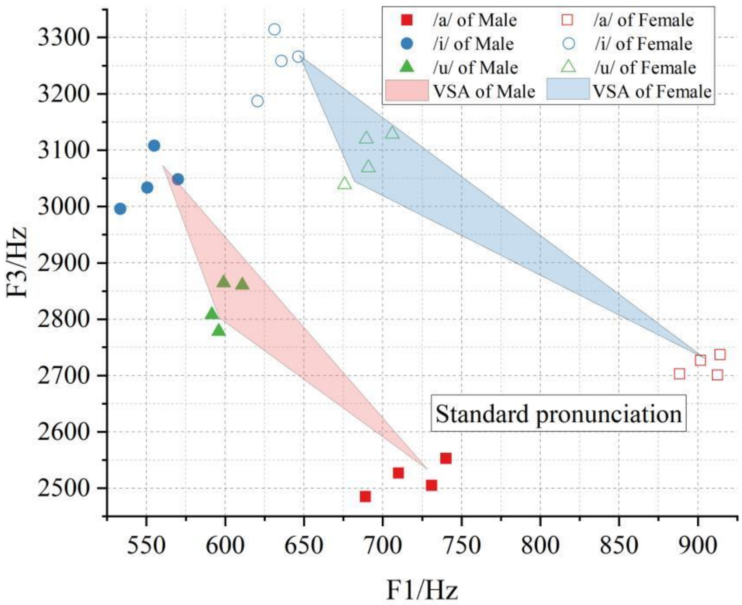


Figure 9: Acoustic vowel diagrams for standard male and female vocalizations

In the standard vocalization state, the three vowels of the male and female voices constitute a smaller area of 15586 and 13669 Hz², respectively, when the vocal tract is regulated efficiently and consistently.

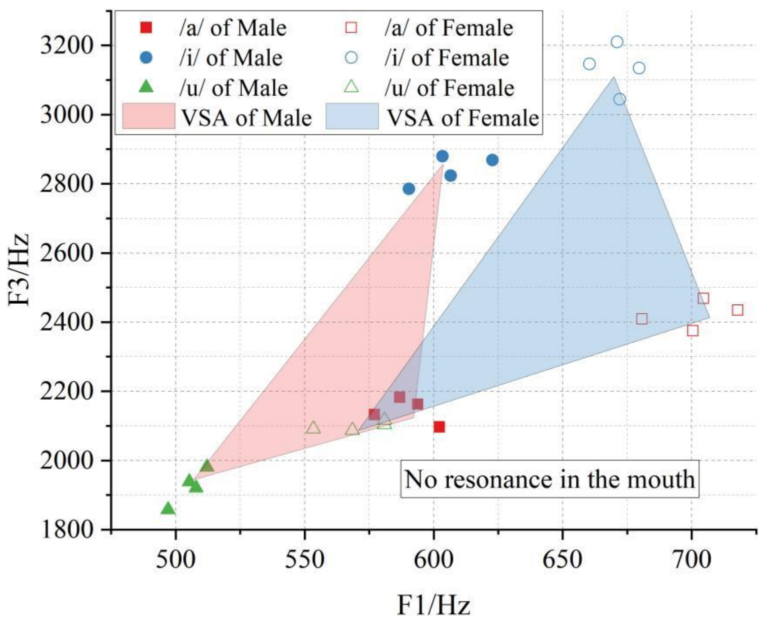


Figure 10: Acoustic vowels without resonance in the mouth cavity

In the oral unresonated cavity, the VSA area was amplified, with male and female voice VSAs reaching 27430 and 41926 Hz², respectively, 4-11 times the norm. It is shown that under oral unresonated cavity conditions alone, vowel acoustic positions are incongruously shifted and the clarity of American singing is reduced.

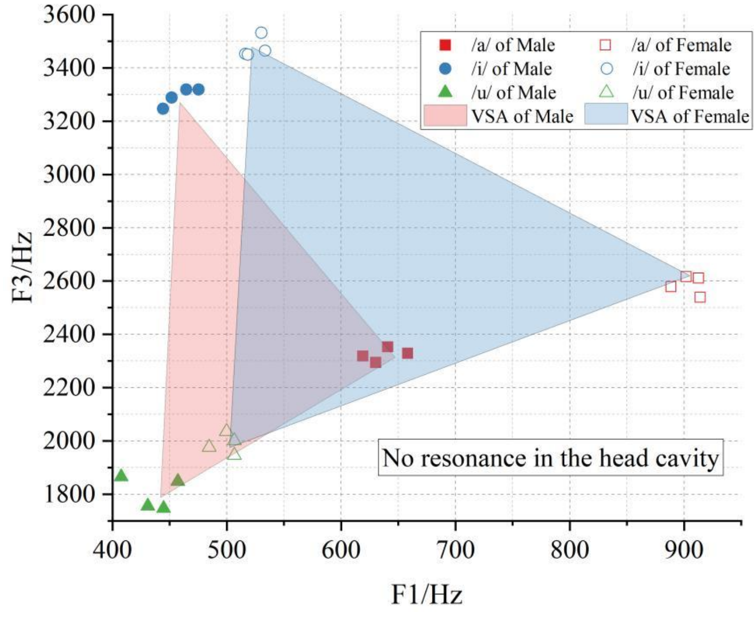


Figure 11: Acoustic vowels without resonance in the head cavity

The most severe loss of control occurs when the headspace is not resonating, with a multiplicative expansion of VSA, especially in the female voice, where VSA = 177,762 Hz², 48 times the norm (VSA = 162,329 Hz² in the male voice). The extreme expansion of area represents a loss of control of vocal tract regulation, and the loss of headspace support results in a drastically out-of-alignment compensatory movement of the singer's tongue and larynx in order to maintain the vocalization, leading to severe distortion of the acoustic targets of the vowels and their discrete separation from each other.

In terms of gender differences, the multiplicative increase in VSA relative to the standard value in both error states was higher in females than in males. It is because the female vocal tract is shorter and the range of fluctuation of their resonance peak frequencies (especially F3) is more sensitive and drastic during resonance misalignment.

In addition, two acoustic spatial indices, FCR (resonance peak concentration rate) and VAI (vowel articulation index), were selected to evaluate the different cavity states.

The FCR reflects the articulation accuracy and sound quality performance by analyzing the distribution and concentration of resonance peaks. The formulas are as follows

$$FCR = \frac{F2/a/ + F2/u/ + F1/i/ + F1/u/}{F2/\mathrm{i}/ + F1/a/} \tag{9}$$

The VAI provides calculation of the ratio of resonance peak frequencies to assess the degree of articulatory clarity and variability.

The values of each indicator for male and female voices in the three states of standard vocalization, cephalic unresonance and oral unresonance are shown in Table 1.

Table 1: Acoustic spatial parameters of male and female voices

		Standard pronunciation	No resonance in the head cavity	No resonance in the oral cavity
VSA	Male	15586±27.39	162329±268.94	27430±118.240
	Female	13669±17.08	177762±278.57	41926±128.62
FCR	Male	1.726±0.263	1.222±0.141	1.511±0.208
	Female	1.682±0.281	1.405±0.177	1.554±0.215
VAI	Male	0.579±0.108	0.818±0.093	0.662±0.114
	Female	0.595±0.097	0.712±0.102	0.644±0.108

The FCR of resonance peak concentration rate of male and female voices during standard vocalization were 1.726 and 1.682, respectively, when the resonance peak energy distribution was concentrated and the sound quality was pure. In contrast, the FCR decreased significantly in both error states, with 1.222 and 1.405, 1.511 and 1.544 for cephalic unresonance and oral unresonance, respectively, and the resonance peaks became dispersed. Meanwhile, the vowel articulation index VAI (which is the reciprocal of FCR) was 0.818 and 0.662 for male voices and 0.712 and 0.644 for female voices in both error states, while both were 0.579 and 0.595 in the standard vocalizations, suggesting that the vowel resonance peaks were more concentrated in the standard vocalizations, and that the sensitivity to the vowels was also higher.

4 Conclusion

The study portrays the core acoustic characteristics of American vocal singing through multi-parameter analysis, and reveals the decisive role of the dynamic regulation of the resonance cavities (head cavity and mouth) in stabilizing the sound quality, focusing the timbre, and maintaining the clarity of vowels. The third resonance peak frequencies of the vowels $/a/$, $/i/$, and $/u/$ dropped by 373, 207, and 903 Hz, respectively, after the male voice lost the support of the head cavity. A more macroscopic acoustic spatial analysis (VSA) shows that the spatial distribution of vowels at this time will also be a sharp expansion of the head and mouth without resonance under the male voice VSA increased to 162329Hz and 27430Hz, respectively, and the female voice VSA is up to nearly fifty times the standard value. Therefore, both the head and mouth cavity resonance is very important to the American vocal singing, and inversely, it can be seen that the singer's excellent American vocal singing is essentially the result of the dynamic synergy of the resonating cavities of the vocal tract.

References

[1] Xie, X. (2021). On the Similarities and Differences Between Bel Canto and Popular Singing. Frontiers, 2(1).

[2] Titze, I. R. (2021). Regulation of laryngeal resistance and maximum power transfer with semi-occluded airway vocalization. The Journal of the Acoustical Society of America, 149(6), 4106-4118.

[3] de Oliveira, K. G. S. C., de Lira, Z. S., da Silva, H. J., Lucena, J. A., & Gomes, A. D. O. C. (2022). Oropharyngeal geometry and the singing voice: immediate effect of two semi-occluded vocal tract exercises. Journal of Voice, 36(4), 523-530.

[4] do Nascimento, G. F., da Silva, H. J., de Oliveira, K. G. S. C., de Lira, S. Z., & Gomes, A. D. O. C. (2022). Relationship Between Oropharyngeal Geometry and Acoustic Parameters in Singers: A Preliminary Study. Journal of Voice.

[5] Herbst, C. T. (2017). A review of singing voice subsystem interactions—toward an extended physiological model of “support”. Journal of voice, 31(2), 249-e13.

[6] Herbst, C. T., & Story, B. H. (2022). Computer simulation of vocal tract resonance tuning strategies with respect to fundamental frequency and voice source spectral slope in singing. The Journal of the Acoustical Society of America, 152(6), 3548-3561.

[7] Berezutskyi, V., & Berezutska, M. (2024). Rare Example of abnormal vocal resonance: a case from Balzac’s novel. Pneumologie, 78(08), 556-560.

[8] Şakalar, A. (2024). The human voice: an organological examination of the voice as an instrument. Türk Organoloji Dergisi, 53-60.

[9] 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.

[10] ŠVEC, J. (2019). Basic Principles and Muscular Adjustments for Fine-tuning Vocal Timbre. The Oxford handbook of singing, 109.

[11] Bradford, Z. (2019). Vocal resonance: Optimising source-filter interactions in voice training. Fusion Journal, (15), 47-70.

[12] Vurma, A. (2022). Amplitude effects of vocal tract resonance adjustments when singing louder. Journal of Voice, 36(2), 292-e11.

[13] Chengqiu, P., & Xi, T. X. (2025). A METHODOLOGICAL FRAMEWORK FOR OPTIMIZING VOCAL RESONANCE: INTEGRATING PHARYNGEAL VOICE

TRAINING AND MODERN ACOUSTICS. In 6 PENANG INTERNATIONAL MULTIDISCIPLINARY CONFERENCE 2025 (6 PIMC 2025) (p. 166).