INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

A Study of Soprano Singing in Light of the Source-filter Interaction

Tokihiko Kaburagi

Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka, 815-8540 Japan
kabu@design.kyushu-u.ac.jp

Abstract

We examined the physical interaction between the voice source system in the larynx and the acoustic filter of the vocal tract. The vocal tract of a soprano was first scanned in three dimensions using magnetic resonance imaging (MRI) while she produced four musical notes with different vowels. These images were used to simulate voice production, including the vibratory motion of the vocal folds and the behavior of glottal airflow. Images for the /i/ vowel were used in the simulation, because a good proximity relationship was found between the fundamental frequency and the first impedance peak of the vocal tract. The simulation results revealed that the fundamental frequency (vibration frequency of the vocal folds) was decreased to a large extent by the source-filter interaction especially when their natural frequency was in the proximity of the impedance peak. In a specific case, this frequency lowering had the effect of changing the acoustic load of the vocal tract exerted on the vocal folds so that their vibratory motion was effectively assisted.
Index Terms: voice production, source-filter interaction, soprano singing

1. Introduction

The mechanisms of speech production can be broadly explained using a linear framework (i.e., source-filter theory), based on the assumption that the voice source system in the larynx and the vocal-tract filter can function independently of each other. However, these two systems are not independent *in vivo* and the vibration frequency of the folds is influenced by the acoustic characteristics of the vocal tract. In particular, the behavior of the vocal folds was found to change significantly when the vibration frequency approaches peak input impedance of the vocal tract [1, 2, 3, 4, 5, 6, 7]. These studies have reported that source-filter interaction makes the vocal-fold vibration unstable, causing a change in voice register, a sudden jump of the fundamental frequency, sub-harmonics of voice source signals, and even suppression of vocal-fold oscillation.

In this study, source-filter interaction was examined using vocal-tract data taken from a soprano singer. In soprano singing, the fundamental frequency reaches 1000 Hz or even higher and the fundamental frequency can be very close to a formant. Another feature of singing, formant tuning, also supports the occurrence of such an interaction. Formant tuning (or resonance tuning) is a phenomenon in which the frequency of a formant is adjusted in the proximity of the fundamental frequency [8, 9, 10]. The range of this vocal-tract adjustment depends on the individual singer, but, on average, the first formant frequency becomes very close to the fundamental frequency when the fundamental frequency is around 600 to 1000 Hz for sopranos.

Because the source-filter interaction can cause voice instability, formant tuning may lead to an unstable state of phonation. However, soprano singers can avoid this risk by adaptively controlling the formant frequency so that the vocal-tract load is inductive. On the other hand, Ishizaka and Flanagan [1] reported that the vibration frequency of the vocal folds can change as a function of the vocal-tract load, especially when the vibration frequency is close to the formant frequency. This finding suggests that, besides voice instability, soprano singers can suffer from difficulties controlling the fundamental frequency under the influence of the source-filter interaction. Therefore, we focused on the possible effects of the source-filter interaction during the act of soprano singing and investigated how the vibration frequency of the vocal folds can be affected by the interaction.

We here observed the shape of the vocal tract using magnetic resonance imaging (MRI) while a soprano produced a range of musical notes. Scanned images were first used to investigate how the participant changed their vocal tracts morphologically and acoustically, depending on the note and type of the vowel. A computer simulation of the voice production process was then performed using models of the vocal folds and the vocal tract. We performed simulation experiments with and without the interaction and compared the results to show how the source-filter interaction can affect the motion of the vocal folds, including their vibratory frequency.

2. Morphological and acoustic analysis of the vocal tract

MRI was used to image the vocal tract of an amateur soprano. Sectional images were scanned in three dimensions while the participant produced four musical notes (A3 [220 Hz], E4 [330 Hz], D5 [587 Hz], and A5 [880 Hz]), with each of the /a/ and /i/ vowels. In addition to these singing tasks, the vocal tract was scanned when the participant spoke the /a/ and /i/ vowels. The shape of the upper and lower teeth was then measured independently, while the participant bit the tongue softly so that the outline of the teeth could be distinguished in the images. Each slice was measured in the sagittal plane. The slice thickness was 3 mm for the vocal tract and 1 mm for the teeth. There were 18 slices for the vocal tract and 88 slices for the teeth. The resolution of each slice was 512×512 pixels covering a scanning area of 256×256 mm^2.

Figure 1 shows the midsagittal images of the vocal tract. For the /a/ vowel, the tongue posture for A3 was the same as that for speech. When the note increased from A3 to E4 and from E4 to D5, the tongue shape for the /a/ vowel changed only very slightly, but it exhibited a significant change when the note increased from D5 to A5. For the /i/ vowel, the tongue moved gradually to the posterior direction as the note became higher. As a result, the tongue posture was relatively similar between both vowels for the highest note [11].

To investigate the input impedance of the vocal tract seen from the glottis, we estimated the cross-sectional area of the vocal tract. The vocal tract was divided along the center line into 40 sections of equal length and the cross-sectional area in

http://dx.doi.org/10.21437/Interspeech.2019-1153

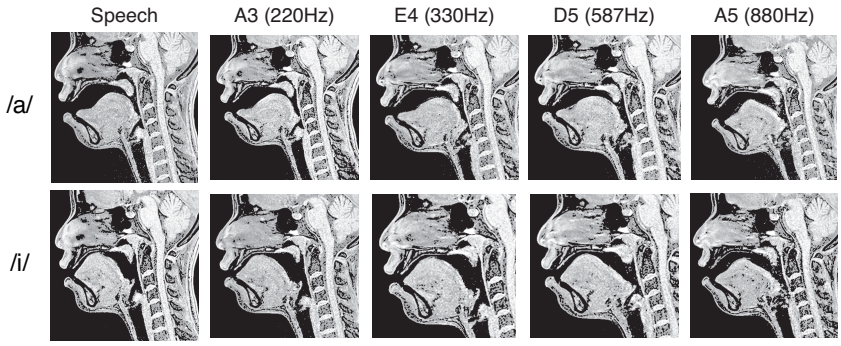

Figure 1: *Midsagittal images of the vocal tract for the /a/ and /i/ vowels measured with MRI. The leftmost image was taken when the participant uttered a vowel. Other images represent vocal-tract shapes when the participant sang musical notes, A3, E4, D5, and A5.*

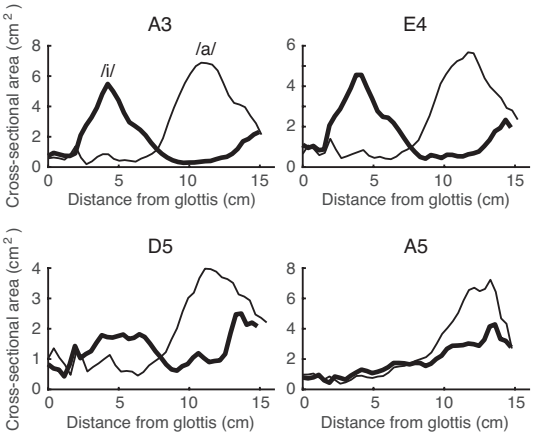

Figure 2: *The thin and thick lines represent the area function for the /a/ and /i/ vowels, respectively.*

the direction perpendicular to the center line was obtained using the vocal-tract images with the teeth superimposed. Figure 2 shows the estimated area function as a function of the distance from the glottis. For notes A3 and E4, a clear difference of the vocal-tract shape can be seen between the /a/ and /i/ vowels. The cross-sectional area of the posterior cavity noticeably decreased for the /i/ vowel when the note changed from E4 to D5 and from D5 to A5, because the tongue moved in the posterior direction. For note A5, the vocal tract had a frontal cavity for both vowels, in accordance with the vocal-tract images in Fig. 1.

The input impedance was then calculated with a frequency-domain acoustic tube model [12]. When the vocal tract sections from the glottis to the lip opening were defined as $1, 2, ..., N$, the propagation matrix is given as

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \prod_{i=1}^{N} \begin{pmatrix} A_i & B_i \\ C_i & D_i \end{pmatrix}. \quad (1)$$

The matrix components, A_i, B_i, C_i, and D_i, are determined by the length, L_i, and cross-sectional area, S_i, for the ith section such that $A_i = \cosh(\sigma L_i/c)$, $B_i = -(\rho c/S_i)\gamma \sinh(\sigma L_i/c)$, $C_i = -(S_i/\rho c)(\sinh(\sigma L_i/c))/\gamma$, and $D_i = \cosh(\sigma L_i/c)$. α, β, γ, and σ are frequency-dependent parameters used in previous studies [12], such that $\alpha = \sqrt{j\omega c_1}$, $\beta = j\omega\omega_0^2/\{(j\omega + a)j\omega + b\} + \alpha$, $\gamma = \sqrt{(\alpha + j\omega)/(\beta + j\omega)}$, and $\sigma = \gamma(\beta + j\omega)$, where $a = 130\pi$ rad/s, $b = (30\pi)^2$ (rad/s)^2, $c_1 = 4$ rad/s, $\omega_0^2 = (406\pi)^2$ (rad/s)^2, and $j = \sqrt{-1}$. ρ is the air density and c is the speed of sound. In this study, ρ was set to 1.184×10^{-3} g/cm^3 and c to 34630 cm/s. The input impedance of the vocal tract, Z, was calculated as follows under the assumption of complete glottal closure:

$$Z = \frac{AZ_p + B}{CZ_p + D}, \quad (2)$$

where Z_p is the radiation impedance of the lip aperture [13].

Figure 3 compares calculated input impedances. The impedance value for the lowest note was markedly different between the /a/ and /i/ vowels. As the note became higher, however, the frequency of the first impedance peak increased and that of the second peak decreased for the /i/ vowel. The impedance characteristics were then relatively similar between the /a/ and /i/ vowels for the highest note. For the /i/ vowel, the magnitude of the first impedance peak increased as the note became higher.

Another important feature was the relationship between the fundamental frequency and the frequency of the first impedance peak. For the /a/ vowel, this peak frequency was almost the same among every note, and the fundamental frequency approached the peak as the note became higher. For the /i/ vowel, the peak frequency increased as the note became higher. In addition, the fundamental frequency was almost the same as the peak frequency, particularly for D5 and A5, indicating the occurrence of the vocal-tract adjustment (i.e., formant tuning). When the formant frequency is higher than the fundamental frequency, the acoustic load of the vocal tract is inductive and the inductive load is preferable for the maintenance of vocal-fold oscillation [4]. The first peak frequency for /a/ was almost stable from A3 to D5 and was sufficiently higher than the fundamental frequency, indicating that the vocal-tract load was inductive. For the /i/ vowel, in contrast, the participant should control the vocal tract so that the peak frequency was higher than the fundamental one to keep the vocal fold vibration stable.

3. Simulation model of the voice production

Next, the effects of the source-filter interaction were examined using a mechanical model of the vocal folds [14] and an acoustic tube model of the vocal tract [12]. We examined the behavior of the vocal fold when the acoustic characteristics of the vocal tract were altered.

By assuming that the air pressure at the separation point is equivalent to the atmospheric pressure, Bernoulli's principle

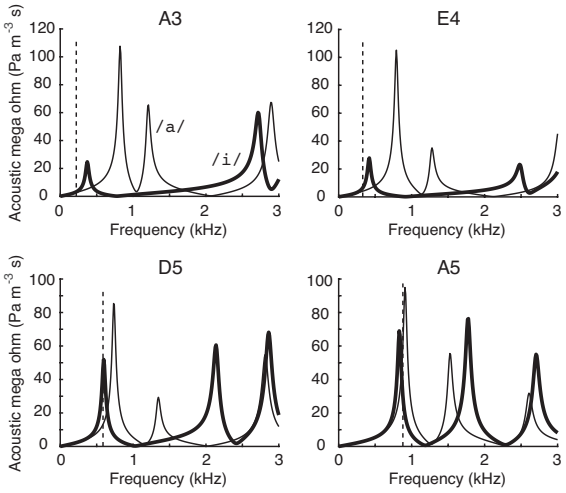

Figure 3: *Input impedance of the vocal tract. The thin and thick lines represent the absolute value of the impedance for the /a/ and /i/ vowels, respectively. The dotted line shows the nominal frequency of each note.*

suggests the following relationship upstream of the separation point:

$$\Delta p = \frac{1}{2}\rho\left(\frac{u_g}{S_s}\right)^2, \tag{3}$$

where Δp is the pressure difference between the glottal entrance and the separation point, u_g is the volume flow that passes through the glottis, and S_s is the opening area of the glottis at the separation point. To represent the source-filter interaction, we consider the acoustic pressure at the entrance of the vocal tract, p_s, in addition to the static tracheal pressure, p_0, as

$$\Delta p = p_0 - p_s. \tag{4}$$

The value of this acoustic pressure can be determined as [12]

$$p_s(t) = Z_0 u_g(t) + \int_0^\infty r(s)\{p_s(t-s) + Z_0 u_g(t-s)\}ds. \tag{5}$$

$r(t)$ is the inverse Fourier transform of the reflection function and $Z_0 = \rho c/S_1$, where S_1 is the cross-sectional area at the entrance of the vocal tract. The reflection function of the vocal tract, R, can be determined from the input impedance, Z, such that $R = (Z - Z_0)/(Z + Z_0)$.

Discretization of Eq. (5) gives

$$p_s(n) = Z_0 u_g(n) + \sum_{k=0}^{K-1} r(k)\{p_s(n-k) + Z_0 u_g(n-k)\}, \tag{6}$$

where n is the time step and K is the effective length of $r(t)$. By combining Eqs. (3), (4), and (6), $u_g(n)$ can be solved as

$$u_g(n) = \frac{-A_2 + \sqrt{A_2^2 + 2A_1(p_0 - A_3)}}{A_1}, \tag{7}$$

where

$$A_1 = \frac{\rho}{S_s^2}, \qquad A_2 = \frac{1 + r(0)}{1 - r(0)} Z_0,$$

and

$$A_3 = \frac{1}{1 - r(0)} \sum_{k=1}^{K-1} r(k)\{p_s(n-k) + Z_0 u_g(n-k)\}.$$

The values of $p_s(n)$ and $u_g(n)$ are thus calculated from Eqs. (6) and (7), respectively, at each time step of the simulation.

Together with the fluid-dynamic and acoustic behavior of the air, a symmetrical two-mass model of the vocal folds [14] was adopted here, because this model is capable of computing essential divergent-convergent changes of the glottal shape. Each vocal fold is constructed by two point masses, m_1 and m_2, respectively located at the lower and upper parts of the fold, and three plates, which connect the inlet and outlet of the glottis and the point masses. Masses are connected to the fixed wall by dampers of resistance r_1 and r_2 and linear springs with Hooke's constants k_1 and k_2. The two masses are joined by another linear spring of constant k_{12}.

The motion of the lower and upper parts of the vocal fold is respectively expressed as

$$m_1\frac{d^2y_1}{dt^2} + r_1\frac{dy_1}{dt} + k_1 y_1 + k_{12}(y_1 - y_2) = f_1 \tag{8}$$

and

$$m_2\frac{d^2y_2}{dt^2} + r_2\frac{dy_2}{dt} + k_2 y_2 + k_{12}(y_2 - y_1) = f_2\,, \tag{9}$$

where y_1 and y_2 are the displacement of the masses. The absolute mass position is $y_1 + y_0$ and $y_2 + y_0$, where y_0 is the common resting position. The effect of collision is incorporated by increasing the values of the mechanical constants of the springs and dampers [14].

f_1 and f_2 are the driving force of each mass, which can be estimated by multiplying the pressure along the axis of the glottal space with the surface area of the vocal fold:

$$f_i = l_g\left\{\int_{x_{i-1}}^{x_i}\frac{x - x_{i-1}}{x_i - x_{i-1}}p(x)dx + \int_{x_i}^{x_{i+1}}\frac{x_{i+1} - x}{x_{i+1} - x_i}p(x)dx\right\}, \tag{10}$$

where $x_i (i = 0, 1, 2, 3)$ represents the position of the glottal inlet, two point masses, and glottal outlet along the glottal midline. l_g is the length of each vocal fold. The pressure, $p(x)$, is the sum of the static pressure from the lungs and the acoustic pressure, p_s.

The geometry of the glottis and the values of the model parameters included in Eqs. (8) and (9) were set following the literature [14]. The length of the vocal folds was 1.4 cm, the depth of the membranous part $(x_2 - x_1)$ was 0.2 cm, and the depth of both the inlet $(x_1 - x_0)$ and outlet $(x_3 - x_2)$ of the glottis was 0.02 cm. The height of the glottal entrance was 1.8 cm. The rest position of the vocal fold masses, y_0, was 1.4×10^{-4} cm. The value of each model parameter was the same for the upper and lower parts of the vocal fold. The mass parameter was $m_1 = m_2 = 0.1$ g and the stiffness parameter was set so that the natural frequency, $\frac{1}{2\pi}\sqrt{\frac{k_1}{m_1}}$ and $\frac{1}{2\pi}\sqrt{\frac{k_2}{m_2}}$, was 100 Hz. The stiffness connecting the upper and lower parts was $k_{12} = 0.6k_1$. The damper parameter was $r_1 = 0.2\sqrt{k_1 m_1}$ and $r_2 = 0.2\sqrt{k_2 m_2}$, respectively. To change the natural frequency of the vocal folds, a control parameter, q, was introduced. The mass parameter was then divided by q and the stiffness parameter was multiplied by q, resulting in the natural frequency of $100q$ Hz. The air density was 1.184×10^{-3} g/cm^3 and the sound velocity was 34630 cm/s. The sampling frequency of the time domain simulation was 20 kHz and the number of points in calculating the input impedance of the vocal tract was 2^{12}.

4. Simulation results

Voice production simulation was first performed to adjust the values of model parameters, q and p_0 in combination, so that the vibration frequency of the vocal folds was close to the fundamental frequency of singing voices during the MRI measurement. Here, the vocal-tract shape for the /i/ vowel was used because a proximity relationship between the fundamental frequency and the impedance peak was observed. The results are summarized in Table 1. The absolute error between the actual and reproduced pitches was less than 4% for every note.

Next, computer simulation was performed for every combination of three vocal fold conditions and four vocal tract conditions. The vocal fold condition was the set of the natural frequency and the lung pressure value shown in Table 1. The vocal tract condition was the input impedance of the /i/ vowel for one of the notes, A3, E4, D5, or A5. The vibration frequency of the vocal folds was then calculated to examine how the behavior of the vocal folds changes depending on the acoustic load of the vocal tract under the influence of the source-filter interaction.

Figure 4 shows the results. The absolute value and the imaginary part of the vocal-tract input impedance was drawn with blue and red lines, respectively. The vibration frequency of the vocal folds is plotted with vertical lines for each vocal fold condition. The red line shows the vibration frequency simulated with the source-filter interaction. The vibration frequency without the interaction is shown as the broken vertical line. To disable the effects of the interaction, the pressure difference was calculated as $\Delta p = p_0$ in Eq. (4) and the pressure $p(x)$ in Eq. (10) was computed only from the static pressure in the glottis. When the interaction was disabled, the vibration frequency was identical irrespective of the vocal-tract condition if the vocal fold condition was the same.

When the natural frequency was 210 Hz, this frequency was lower than the first impedance peak for every vocal tract condition. The vibration frequency of the vocal folds was almost the same as that without the interaction. In addition, the vibration frequency did not substantially change among the four vocal tract conditions, indicating that the vocal fold motion was less affected by the change in the input impedance of the vocal tract.

The natural frequency of 340 Hz was very close to the first impedance peak for A3. Among the four notes, the magnitude of the impedance at 340 Hz was the largest for this note. Moreover, the vibration frequency with the interaction was noticeably lower than the frequency without the interaction, suggesting the occurrence of a strong source-filter interaction. For E4 and A5, the frequency with the interaction was lower than that without the interaction to a small degree.

The natural frequency of 630 Hz was higher than the frequency of the first impedance peak for A3, E4, and D5. For

Table 1: *Parameter values for reproducing the fundamental frequency of the participant's singing voice. "Nom" is the nominal pitch, "Act" is the actual pitch produced by the participant. "Nat" is the natural frequency of the vocal fold model, and "Rep" is the reproduced pitch by the simulation.*

Note	Pitch (Hz)				Lung pressure (cmH_2O)
	Nom	Act	Nat	Rep	
A3	220	223	210	231	10
E4	330	337	340	347	15
D5	587	587	630	588	25

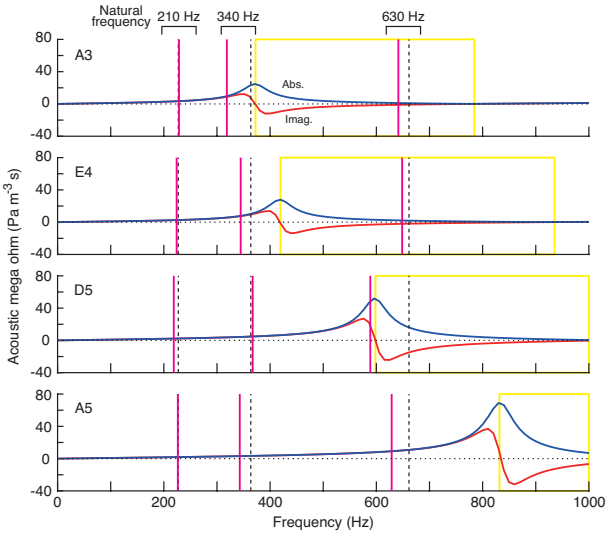

Figure 4: *Results of the voice production simulation for the vocal-tract configurations of the /i/ vowel sung at the notes of A3, E4, D5, and A5. The blue line shows the absolute value of the input impedance and the red line shows its reactance. The yellow box shows the frequency region in which the reactance is negative and the vocal-tract load is capacitive. Outside this region, the load is inductive. Vibration frequency of the vocal folds with and without the source-filter interaction are drawn by solid and broken vertical lines, respectively.*

A3 and E4, this natural frequency was much higher than the impedance peak. The results revealed that the change in the vibration frequency with and without the interaction was relatively small for these notes. For D5, the natural frequency was higher than the negative peak of the imaginary part of the impedance to a small extent, indicating that the acoustic load of the vocal tract was capacitive. As a result, the vibration frequency was much lower than the natural frequency under the interaction. At that vibration frequency, the imaginary part of the impedance was positive and the acoustic load was inductive, which is suited to the maintenance of the vocal fold vibration.

5. Conclusions

The change in vibration frequency of the vocal folds was examined using computer simulation when the interaction was switched on and off. This change was significant when the natural frequency of the folds was in the proximity of an impedance peak of the vocal tract, and the interaction generally decreased the vibration frequency. In particular, when the vocal tract condition was D5 and the natural frequency was 630 Hz, the acoustic load of the vocal tract was capacitive at the vibration frequency without the interaction. However, when the interaction was switched on, the load was inductive as a result of the lowering of the vibration frequency. As such, our findings imply that the fundamental frequency may be influenced by the source-filter interaction in addition to the physiological adjustment of the vocal folds. Fine adjustment of the fundamental frequency may become difficult under a strong interaction, and it requires the simultaneous control of the larynx and the vocal tract. Further study is needed for exploring such synergistic control of the organs when an experienced soprano adjusts the fundamental frequency while managing undesirable interaction effects.

6. References

[1] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J*, vol. 51, pp. 1233–1268, 1972.

[2] M. Zañartu, L. Mongeau, and G. R. Wodicka, "Influence of acoustic loading on an effective single mass model of the vocal folds," *J. Acoust. Soc. Am.*, vol. 121, pp. 1119–1129, 2007.

[3] I. T. Tokuda, J. Horáček, J. G. Švec, and H. Herzel, "Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments," *J. Acoust. Soc. Am.*, vol. 122, pp. 519–531, 2007.

[4] I. R. Titze, "Nonlinear source-filter coupling in phonation: Theory," *J. Acoust. Soc. Am.*, vol. 123, pp. 2733–2749, 2008.

[5] T. Kaburagi, "Voice production model integrating boundary-layer analysis of glottal flow and source-filter coupling," *J. Acoust. Soc. Am.*, vol. 129, pp. 1554–1567, 2011.

[6] M. Zañartu, D. D. Mehta, J. C. Ho, G. R. Wodicka and R. E. Hillman, "Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study," *J. Acoust. Soc. Am.*, vol. 129, pp. 326–339, 2011.

[7] Y. Uezu and T. Kaburagi, "A measurement study on voice instabilities during modal-falsetto register transition," *Acoust. Sci. & Tech.*, vol. 37, pp. 267–276, 2016.

[8] J. Sundberg, "Formant technique in a professional female singer," *Acustica*, vol. 32, pp. 89–96, 1975.

[9] E. Joliveau, J. Smith, and J. Wolfe, "Vocal tract resonances in singing: The soprano voice," *J. Acoust. Soc. Am.*, vol. 116, pp. 2434–2439, 2004.

[10] M. Garnier, N. Henrich, J. Smith, and J. Wolfe, "Vocal tract adjustments in the high soprano range," *J. Acoust. Soc. Am.*, vol. 127, pp. 3771–3780, 2010.

[11] C. Johansson, J. Sundberg, and H. Wilbrand, "X-ray study of articulation and formant frequencies in two female singers," *STL-QPSR*, vol. 23, pp. 117–134, 1982.

[12] M. M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Acoust. Speech Signal Proc., vol. 35, pp. 955–967, 1987.

[13] J. L. Flanagan, *Speech Analysis Synthesis and Perception (2nd. Ed.)* (Springer Verlag, New York, 1972), pp. 36–38.

[14] N. J. C. Lous, G. C. Hofmans, R. N. J. Veldhuis, and A. Hirschberg, "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design," *Acust. Acta Acust.*, vol. 84, pp. 1135–1150, 1998.