九州大学学術情報リポジトリ
Kyushu University Institutional Repository

Source-filter interaction in phonation: A study using vocal-tract data of a soprano singer

Kaburagi, Tokihiko
Faculty of Design, Kyushu University

Ando, Momoyo
Graduate School of Design, Kyushu University

Uezu, Yasufumi
Human Information Science Laboratory, NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation

https://hdl.handle.net/2324/7178832

出版情報：Acoustical Science and Technology. 40 (5), pp.313-324, 2019. 日本音響学会
バージョン：
権利関係：© 2019 by The Acoustical Society of Japan

KYUSHU UNIVERSITY

Acoust. Sci. & Tech. **40**, 5 (2019) ©2019 The Acoustical Society of Japan

PAPER

Source-filter interaction in phonation: A study using vocal-tract data of a soprano singer

Tokihiko Kaburagi[1,*], Momoyo Ando[2] and Yasufumi Uezu[3]

[1]*Faculty of Design, Kyushu University,*
4–9–1 Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan
[2]*Graduate School of Design, Kyushu University,*
4–9–1 Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan
[3]*Human Information Science Laboratory, NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, 3–1, Morinosato Wakamiya, Atsugi, 243–0198 Japan*

(*Received 9 August 2018, Accepted for publication 9 May 2019*)

Abstract: We simulated acoustic 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 while she produced four musical notes (A3, E4, D5, and A5) with /a/ and /i/. 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 interaction especially when their natural frequency was in the proximity of the impedance peak. In a specific case, the acoustic load of the vocal tract exerted on the vocal folds changed as a result of the interaction, so the vibratory motion was effectively assisted. These interaction effects were also examined in terms of the phase relation among the temporal waveforms of the glottal variables.

Keywords: Phonation, Glottal sound source, Source-filter interaction, MRI, Soprano singing

PACS number: 43.70.Bk, 43.70.Gr [doi:10.1250/ast.40.313]

1. INTRODUCTION

The mechanisms of speech production can be broadly explained using a linear framework (i.e., source-filter theory) [1], 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 acoustic systems are not independent in vivo. In a classic study of the two-mass vocal fold model [2], Ishizaka and Flanagan demonstrated that 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.

Since Ishizaka and Flanagan's study, the dependence between the voice source system and the vocal-tract filter (i.e., source-filter interaction) have been examined extensively [3–6]. Titze [5] conducted a computer simulation of the voice production process when the fundamental frequency changes in time such that modal and falsetto registers are switched via interaction. The study demonstrated that the effect of interaction was more accentuated when the cross-sectional area of the epilarynx tube is narrow, and accentuated interaction induced a large pitch jump during the register transition.

Titze [5] and Titze and Worley [7] interpreted the interaction more specifically in terms of the reactance of input impedances arising from the acoustic load of the sub- and supra-glottal tracts, because the phase response of these impedances provides information about the relative timing between flow and pressure waveforms in the vicinity of the glottis. When the acoustic load of the vocal tract is inductive, for example, positive pressure results when the glottis is opening and the flow is increasing because of the phase advance properties of the load. This positive pressure pushes the left and right vocal folds and assists the opening behavior of the glottis. The phase response therefore determines the quality of the source-

*e-mail: kabu@design.kyushu-u.ac.jp

filter interaction. The magnitude response, on the other hand, gives an indication of the quantity and strength of the interaction, since it represents the amplitude of the acoustic pressure fed back from the tracts.

Thus, the acoustic load of the vocal tract can affect the voice-source system, and this interaction has been confirmed not only by computer simulations, but also voice measurements using human subjects [8–10]. These measurement 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. It has also been found that the interaction tends to occur for vowels with a relatively low first formant frequency, such as the /i/ vowel, because the fundamental frequency can readily approach the formant.

In the current study, source-filter interaction was examined using vocal-tract data taken from a soprano singer. In soprano singing, the fundamental frequency reaches 1,000 Hz or even higher. This singing frequency is high enough to cover the first formant frequency of American English vowels uttered by adult female speakers [11]. Therefore, the fundamental frequency can be very close to a formant during the act of soprano singing, resulting in source-filter interaction.

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 certain formant is adjusted in the proximity of the fundamental frequency [12–14] by controlling the vocal tract. 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 1,000 Hz for sopranos. In addition, it has been reported that the formant frequency was on average consistently slightly higher than the fundamental frequency [13]. Note that the acoustic load of the vocal tract is inductive for such conditions, and the inductive load is preferable for the maintenance of vocal-fold oscillation [5].

Because the source-filter interaction can cause voice instability, as described above, 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. In addition, when the fundamental frequency or its harmonic component is in the proximity of the formant, the glottal sound source will be enhanced with ease by the corresponding vocal tract resonance. On the other hand, Ishizaka and Flanagan [2] 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 observed the shape of the vocal tract in three dimensions using magnetic resonance imaging (MRI) while a soprano produced a range of musical notes. MRI is a safe, noninvasive imaging technique for use in humans [15,16]. Previous studies performed volumetric scanning of the vocal tract, reporting that the cross-sectional area function can be accurately estimated [17,18]. Area function data are useful for examining the acoustic characteristics of the vocal tract [19].

In the current study, 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. The acoustic characteristics of the vocal tract were realistic in our simulation study because they were determined from scanned images. In addition, if the participant used formant tuning, the fundamental frequency would be expected to be in the proximity of a peak of the input impedance. We were able to manipulate the mechanical properties of the vocal folds and other phonation conditions, such as the lung pressure, in the computer simulation. The effect of the interaction was also switched on and off in each condition. 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.

The current paper is organized as follows. Section 2 provides a description of the image acquisition technique and results. The method for calculating the input impedance is also described. The method of computer simulation is described in Sect. 3, and the results are presented in Sect. 4. Section 5 provides a discussion of our findings and the conclusions that can be drawn from the current results.

2. MORPHOLOGICAL AND ACOUSTIC ANALYSIS OF THE VOCAL TRACT

2.1. Imaging of the Vocal Tract

MRI was used to non-invasively image the vocal tract. The participant was an amateur soprano studying music and singing at a university while training to become a music teacher. Sectional images of the vocal tract were scanned in three dimensions while the participant produced

each of 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 the supine position. The participant was instructed to sing these notes at mezzo forte without using vibrato, and to maintain the configuration of the vocal tract with as much stability as possible. The guide tone was presented to the participant for each note just before the scanning was started. In addition to these singing tasks, the vocal tract was scanned when the same participant spoke the /a/ and /i/ vowels at a comfortable pitch and loudness. The shape of the upper and lower teeth was then measured independently, because they are as transparent as air in standard vocal-tract scanning [18]. The oral shape was scanned using the usual scanning procedures, 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 $256 \times 256\,\mathrm{mm}^2$. The measurement was made by the ATR-Promotions Inc. (Kyoto, Japan).

Figure 1 shows the midsagittal images of the vocal tract for speech and singing. The main speech articulators, including the tongue, lips, and soft palate, can be clearly seen in the images. For the /a/ vowel, the tongue posture obtained for the A3 note 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 [20].

2.2. Input Impedance of the Vocal Tract

To investigate the input impedance of the vocal tract seen from the glottis, we estimated the cross-sectional area of the vocal tract from each set of volumetric images. The images for the upper and lower teeth were first superimposed on the vocal-tract images, based on an affine congruent transformation [18]. Before applying the transformation, the position of the teeth was manually marked on images. When the transformation was performed and the marked teeth area was mapped onto the tract images, the correlation coefficient between the teeth and tract images was calculated for the marked area. The optimal affine transformation was then determined so that the value of the correlation coefficient was greatest.

The area function was obtained using the vocal-tract images with the teeth superimposed. First, the inferior and superior outlines of the vocal tract were traced by sight on the midsagittal plane. The center line of the tract, representing the axis of wave propagation, was determined so that the distances of each node point on the center line from the two outlines were equal (see Fig. 2). The vocal tract was then divided along the center line into 40 sections of equal length. Finally, the cross-sectional area in the direction perpendicular to the center line was obtained for each section.

Figure 2 shows the estimated area function for four musical notes 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 figure shows that the vocal tract had a frontal cavity both for the /a/ and /i/ vowels. This finding is in accordance with the features of the vocal-tract images shown in Fig. 1.

The input impedance was calculated from cross-sectional area function data with a frequency-domain acoustic

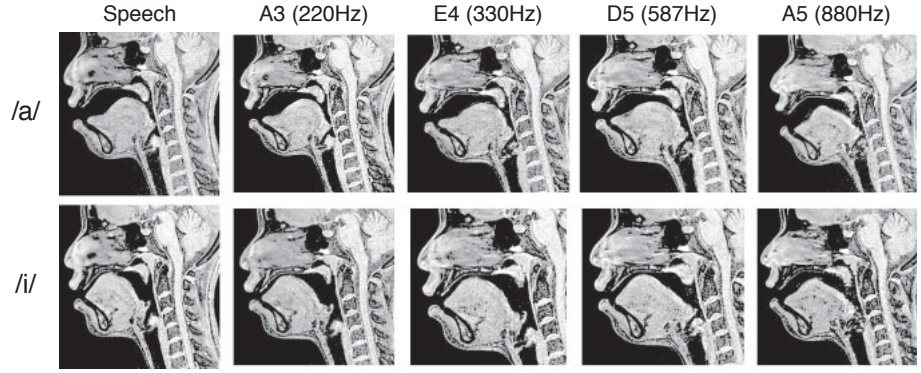

Fig. 1 Midsagittal images of the vocal tract for the /a/ and /i/ vowels measured with MRI. The leftmost image was taken when the participant uttered a vowel. Other images represent vocal-tract shapes when the participant sang four musical notes, A3, E4, D5, and A5.

Fig. 2 Illustration of the method for determining the center line of the vocal tract (left) and estimated cross-sectional area function of the vocal tract for four musical notes. The thin and thick lines represent the area function for the /a/ and /i/ vowels, respectively.

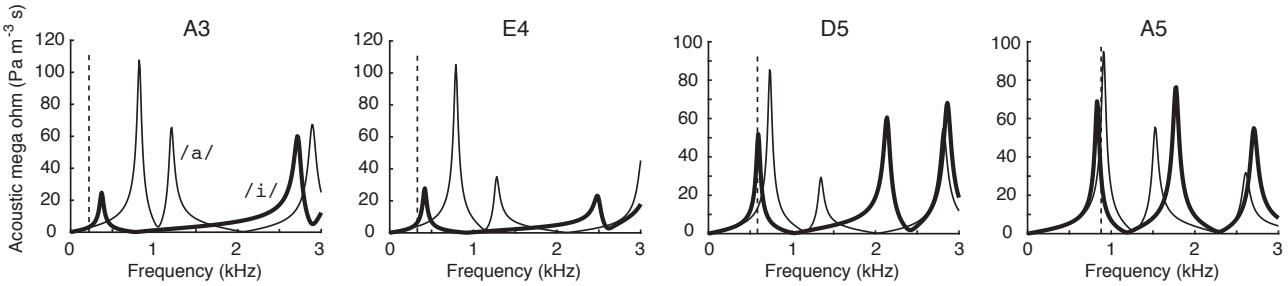

Fig. 3 Input impedance of the vocal tract calculated for four musical notes. The thin and thick lines represent the absolute value of the impedance for the /a/ and /i/ vowels, respectively. The dotted line shows the nominal frequency of each note, namely, 220 Hz for A3, 330 Hz for E4, 587 Hz for D5, and 880 Hz for A5. The actual fundamental frequency produced during the MRI measurement was not always in agreement with the nominal frequency.

tube model [19], which properly accounts for energy losses along the wall of the vocal-tract tube. When the vocal tract sections from the glottis to the lip opening were defined as $1, 2, \ldots, N$, the propagation matrix is given as

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

Here, 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 [19], 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 34,630 cm/s.

The transfer function, H, and input impedance, Z, of the vocal tract were calculated as follows under the assumption of complete glottal closure:

$$H = \frac{1}{CZ_p + D} \tag{2}$$

and

$$Z = \frac{AZ_p + B}{CZ_p + D}, \tag{3}$$

where Z_p is the radiation impedance of the lip aperture [21]. The transfer function is the volume velocity at the lip end over the volume velocity at the glottis. The input impedance is the ratio between the acoustic pressure and the volume velocity at the glottis when the vocal tract is seen from the glottis.

Figure 3 compares calculated input impedances for the /a/ and /i/ vowels. The impedance value for the lowest note was markedly different between the 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. Perceptually, vowel intelligibility was lost for this note and it was difficult to distinguish both vowels. For the /i/ vowel, it should be noted that 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 figure shows that 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) [12–14].

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 [5]. 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 and the participant could sing without changing the vocal tract. For the /i/ vowel, in contrast, the figure suggests that 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.

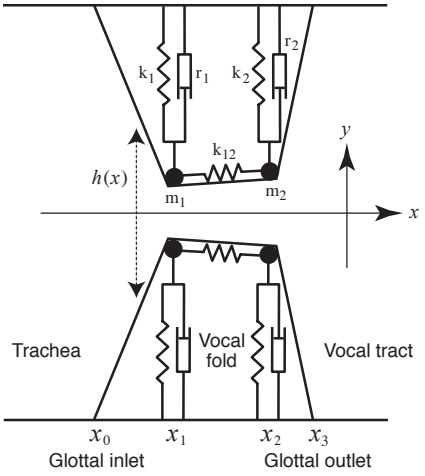

Fig. 4 Construction of the mechanical model of the vocal folds (adapted from [22]).

3. VOICE PRODUCTION SIMULATION

Next, the effects of the source-filter interaction were examined using a mechanical model of the vocal folds [22] and an acoustic tube model of the vocal tract [19]. The acoustic characteristics of the vocal tract were computed from the image data, as explained in the previous section. We examined the behavior of the vocal fold when the acoustic characteristics of the vocal tract were altered, as described below.

3.1. Simulation Model

The physical quantity of glottal sound source signals is the time-varying pattern of airflow that gushes from the glottis. When the vocal folds vibrate to open and close the glottis periodically, the opening area of the glottis changes with time in synchronization with the vocal fold motion. The temporal pattern of the glottal volume flow is thus determined by the vocal fold motion. When the flow enters the glottal space, the flow streamline is parallel with the wall of the channel. By assuming that the flow is inviscid, steady, and incompressible, the flow behavior can be described by Bernoulli's principle in this laminar flow region. Further downstream, the viscosity of air can no longer be ignored and a phenomenon called flow separation takes place. The separation point is usually in the vicinity of the outlet of the glottal space, but the point is known to change and move depending on the shape of the glottal space [23,24].

By assuming that the air pressure at the separation point is equivalent to the atmospheric pressure, Bernoulli's principle suggests the following relationship upstream of the separation point:

$$\Delta p = \frac{1}{2}\rho\left(\frac{u_g}{S_s}\right)^2, \tag{4}$$

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.

This relationship suggests that u_g is mainly determined by the static pressure in the trachea and the cross-sectional area of the glottal space at the flow 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 lung pressure, p_0, as

$$\Delta p = p_0 - p_s. \tag{5}$$

The value of this acoustic pressure at time t can be determined as [19]

$$p_s(t) = Z_0 u_g(t) + \int_0^\infty r(s)\{p_s(t-s) + Z_0 u_g(t-s)\}ds. \tag{6}$$

$r(t)$ 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

$$R = \frac{Z - Z_0}{Z + Z_0}. \tag{7}$$

Discretization of Eq. (6) gives

$$p_s(n) = Z_0 u_g(n) + \sum_{k=0}^{K-1} r(k)\{p_s(n-k) + Z_0 u_g(n-k)\}, \tag{8}$$

where n is the time step and K is the effective length of $r(t)$. By combining Eqs. (4), (5), and (8), $u_g(n)$ can be solved as

$$u_g(n) = \frac{-A_2 + \sqrt{A_2^2 + 2A_1(p_0 - A_3)}}{A_1}, \tag{9}$$

where

$$A_1 = \frac{\rho}{S_s^2},$$

$$A_2 = \frac{1 + r(0)}{1 - r(0)} Z_0,$$

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. (8) and (9), respectively, at each time step of the simulation.

Together with the fluid-dynamic and acoustic behavior of the air, the mechanical behavior of the elastic body (i.e., the vocal folds) should be incorporated. A symmetrical two-mass model [22] was adopted here, because this model is capable of computing essential divergent-convergent changes of the glottal shape during each oscillatory cycle. According to this divergent-convergent change, the flow separation point can vary so that the vocal fold oscillation is effectively maintained [23,24]. 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} (see Fig. 4).

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

$$m_1 \frac{d^2 y_1}{dt^2} + r_1 \frac{dy_1}{dt} + k_1 y_1 + k_{12}(y_1 - y_2) = f_1 \quad (10)$$

and

$$m_2 \frac{d^2 y_2}{dt^2} + r_2 \frac{dy_2}{dt} + k_2 y_2 + k_{12}(y_2 - y_1) = f_2, \quad (11)$$

where y_1 and y_2 are the displacement of the masses perpendicular to the midline of the glottis. The absolute mass position is $y_1 + y_0$ and $y_2 + y_0$, where y_0 is the common resting position. The stiffness of the springs is linear, even when the vocal folds collide with each other. The effect of collision is incorporated by increasing the values of the mechanical constants of the springs and dampers [22].

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:

$$f_i = l_g \left\{ \int_{x_{i-1}}^{x_i} \frac{x - x_{i-1}}{x_i - x_{i-1}} p(x) dx + \int_{x_i}^{x_{i+1}} \frac{x_{i+1} - x}{x_{i+1} - x_i} p(x) dx \right\}, \quad (12)$$

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, given in Eq. (8). Note that the static pressure in the glottis depends on the glottal shape and is a function of x. The value of $p(x)$ can be determined using Bernoulli's equation upstream of the separation point, and can be set to p_s downstream of the separation point, in accordance with a previously reported method [22]. In addition, integration in Eq. (12) can be performed easily because of the polygonal representation of the glottal shape. The flow separation point is determined using the separation constant [22].

3.2. Simulation Method

Simulation experiments were performed, as described below. The masses of the vocal-fold model were initially set to their resting positions. The acoustic transmission matrices and input impedances of the vocal tract were calculated from specific area function data. Next, the volume flow through the glottis and the pressure distribution along the vocal fold were determined. The driving force of the vocal folds was calculated afterward, and the mechanical equations were solved using an explicit Runge-Kutta method. This procedure was repeated over the desired time steps.

The geometry of the glottis and the values of the model parameters included in Eqs. (10) and (11) were set following the literature [22]. 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 separation constant was 1.2. The air density was 1.184×10^{-3} g/cm^3 and the sound velocity was 34,630 cm/s. The sampling frequency of the time domain simulation was 20 kHz and the number of points in calculating the transfer function, H, and the input impedance, Z, of the vocal tract was 2^{12}.

Table 1 Parameter values for reproducing the fundamental frequency of the participant's singing voice. "Nom." is the nominal pitch presented to the participant in the MRI measurement as a guide tone. "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 (vibration frequency of the folds) obtained 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

4. SIMULATION RESULTS

4.1. Adjustment of the Glottal Parameters

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 was constructed from MRI data obtained for the /i/ vowel. This process was undertaken because, as shown in Fig. 3, the frequency of the first impedance peak increased as the note became higher and a proximity relationship between the fundamental frequency and the impedance peak was observed. This was fitted for examining the effect of the source-filter interaction.

The results of this preliminary experiment are summarized in Table 1. Each simulation was performed for a duration of 0.4 s, and the reproduced fundamental frequency was calculated from a steady portion of the synthesized glottal flow signal using a pitch detection algorithm [25]. The absolute error between the actual and reproduced pitches was less than 4% for every note.

4.2. Effect of the Source-filter Interaction on the Vibration Frequency of the Vocal Folds

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 for one of the notes, A3, E4, or D5. 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 (input impedance) of the vocal tract under the influence of the source-filter interaction.

Figure 5 shows the results, where the vocal tract condition was A3, E4, D5, and A5 from top to bottom. The absolute value and the imaginary part of the vocal-tract input impedance was drawn for each note 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. We also obtained the vibration frequency without the interaction, and the result 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. (5). In addition, the pressure $p(x)$ in Eq. (12) was computed only from the static pressure in the glottis. Note that when the interaction was disabled, the vibration frequency was identical irrespective of the vocal-tract condition if the vocal fold condition was the same.

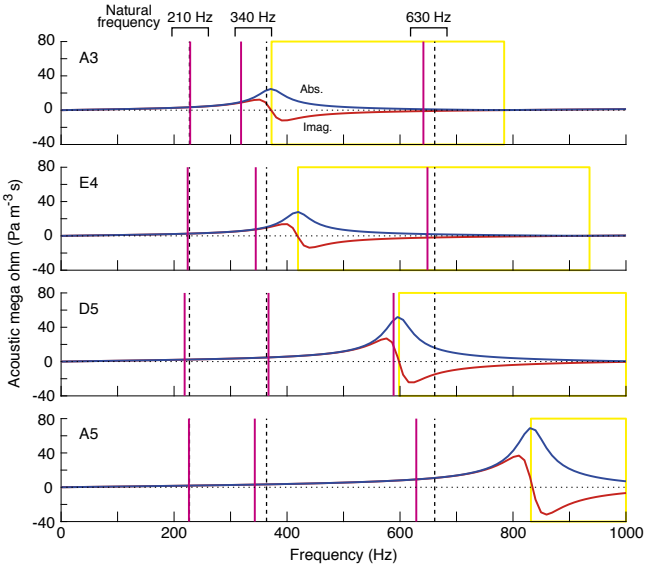

Fig. 5 Results of the voice production simulation. Vibration frequency of the vocal folds with and without the source-filter interaction are drawn by solid and broken vertical lines, respectively. The vocal-tract configurations of the /i/ vowel sung at the notes of A3, E4, D5, and A5 were used in the simulation. The blue line shows the absolute value of the vocal-tract 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.

When the natural frequency was 210 Hz, this natural 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 obtained 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 the note of 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 notes 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 notes A3, E4, and D5. For 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 note 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].

For the vocal-tract condition of A5, the natural frequencies of 340 Hz and 630 Hz were away from the impedance peak, but the figure shows that the vibration frequency decreased with the interaction. This was because the impedance value was especially great for A5 around the peak (see Fig. 3), and the impedance was still large enough at the frequency of 630 Hz. For the natural frequency of 340 Hz, the impedance value was noticeable at the double frequency. These factors were responsible for the interaction and lowering of the vibration frequency.

To examine in more detail the lowering of the vibration frequency observed for the natural frequency of 340 Hz and the vocal tract condition of A3, the vibration frequency of the vocal folds is plotted as a function of their natural frequency in Fig. 6. When the interaction was switched off, the vibration frequency increased almost linearly as the natural frequency increased from 240 Hz to 340 Hz. When the interaction was switched on, the increase in the vibration frequency diminished when the natural frequency approached the first peak (approximately 371 Hz) of the vocal-tract impedance. As a result, the lowering of the vibration frequency caused by incorporating the interaction was significant when the natural frequency approached the impedance peak. In this simulation experiment, the subglottal pressure was a linear function of the natural frequency, set so that the pressure was 10 cmH_2O when the natural frequency was 210 Hz and 15 cmH_2O when the natural frequency was 340 Hz.

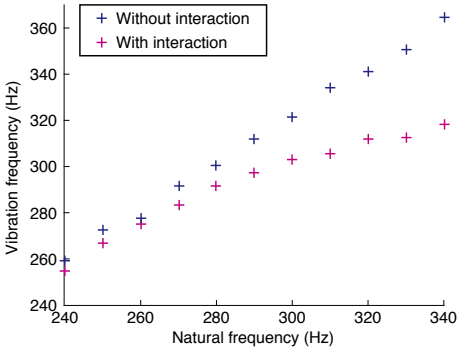

Fig. 6 Vibration frequency of the vocal folds as a function of their natural frequency. The red and blue markers respectively represent the vibration frequency with and without the interaction. The vocal tract condition was A3.

4.3. Effect of Source-filter Interaction on the Waveform of the Glottal Variables

Figure 7 shows simulated waveforms of a range of glottal variables, including acoustic pressure at the entrance of the vocal tract, glottal volume flow, and displacements of the upper and lower vocal-fold masses. The vocal-tract condition was the /i/ vowel sung at the note of A3. For the natural frequency of 210 Hz, the vibration frequency of the vocal folds was almost identical with and without the interaction (top plot of Fig. 5). In the top plot of Fig. 7, the temporal pattern of the volume flow or mass displacement was relatively stable without the interaction. Moreover, it was also relatively stable with the interaction, although a slight perturbation could be observed in the peak value of the volume flow for each pitch period. The standard deviation of the peak value was 5.4 cm^3/s with the interaction, while it was zero without the interaction.

When the natural frequency was 340 Hz and the interaction was not considered, Fig. 7 shows that the instant of each positive peak of the pressure waveform was in accord with that of the flow or mass displacement as indicated by (a) in the figure. This phase relation prevented the closing motion of the vocal folds and hindered their vibration if the interaction was activated while holding this phase relation. However, when the interaction was considered, the vibration frequency of the vocal folds decreased significantly, as shown in Fig. 5. As shown in Fig. 7, the phase relation among the glottal variables changed as a consequence. With interaction, the vocal folds began to open when the pressure had a positive peak as indicated by (b) in the figure. This phase relation was suited for maintaining their vibratory motion.

Finally, at the natural frequency of 630 Hz, the acoustic load of the vocal tract was capacitive. Although the vibration frequency of the vocal folds was different with and without the interaction (Fig. 5), the phase relation among the glottal variables in Fig. 7 appeared similar, where positive peaks of the pressure waveform were observed when the glottis was closed as indicated by (c) in

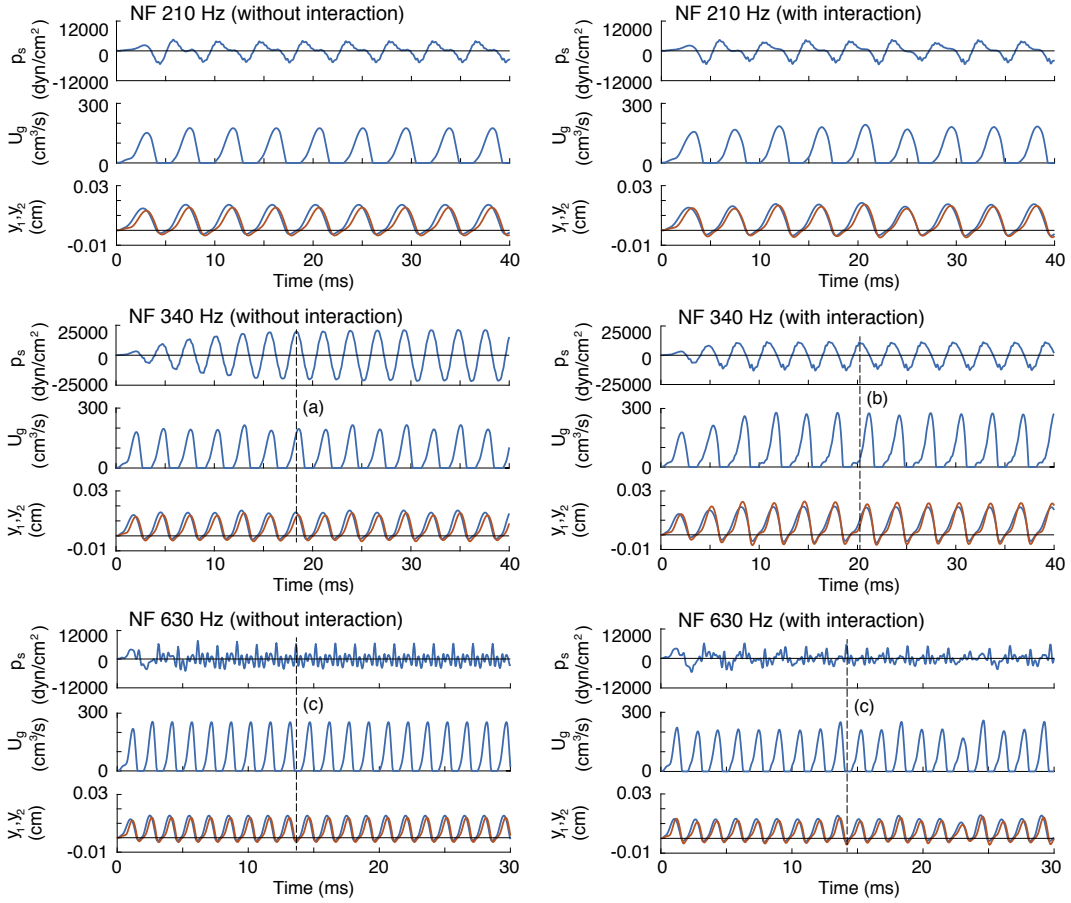

Fig. 7 Simulation results were plotted for the vocal tract condition of A3 in terms of the temporal waveforms of the acoustic pressure at the entrance of the vocal tract (p_s), glottal volume flow (u_g), and displacement of the masses (y_1 and y_2). The blue line represents displacement for the lower mass. The red line represents displacement for the upper mass. The natural frequency of the vocal folds was 210 Hz (top), 340 Hz (middle), and 630 Hz (bottom). In addition, simulation was performed with (right plots) and without (left plots) the source-filter interaction. Broken lines with letters (a) through (c) show the type of phase relation among the glottal variables (see text).

the figure. This phase relation was not suited for the vocal fold vibration, but the lung pressure was 25 cmH_2O (24,517 dyn/cm^2) in the simulation and was much greater than the acoustic pressure p_s. The vocal fold vibration was maintained under the capacitive vocal-tract load because of the high lung pressure. The acoustic pressure waveform comprised high-frequency components for both the with and without conditions, because the frequencies for the fourth harmonic component of the glottal flow were 2,646 Hz and 2,592 Hz for these conditions, and this harmonic component was enhanced by the second impedance peak of 2,715 Hz (see Fig. 3).

Next, Fig. 8 shows the simulation results when the vocal tract condition was D5. The vocal tract load was capacitive at the natural frequency of 630 Hz (Fig. 5), but the vibration frequency decreased significantly with the interaction, and the phase relation among the glottal variables became different between the with and without conditions. The pressure waveform had positive peaks when the glottis was closing without the interaction as indicated by (d) in the figure, whereas it had positive peaks when the glottis was opening with the interaction as indicated by (e). The phase relation was matched for phonation in the interactive condition.

Note that simulated pressure value was much greater in the D5 condition compared with the A3 condition, because the vibration frequency in the D5 condition was very close to the first impedance peak when the interaction was considered (Fig. 5). The maximum pressure value was almost equivalent to the lung pressure (24,517 dyn/cm^2) and the effect of the interaction was significant. Figure 8 also shows that the temporal patterns of the mass displacement and the glottal flow were unstable with the interaction. Regarding the perturbation of the peak value of the volume flow, the standard deviation was approximately 35 cm^3/s with the interaction, while it was zero without the interaction.

5. SUMMARY AND CONCLUSION

The effect of the source-filter interaction was examined

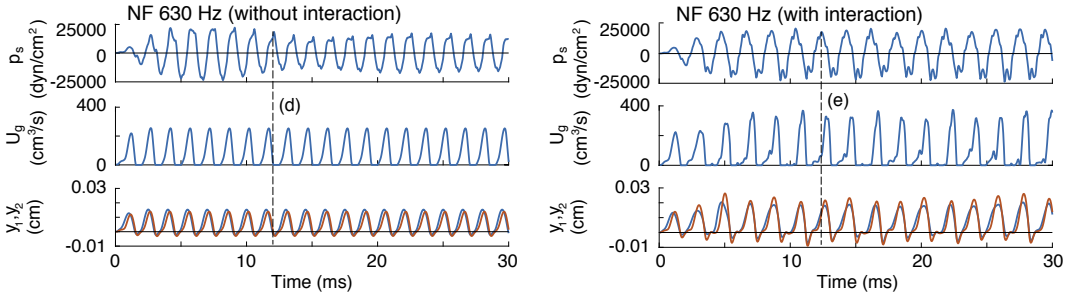

Fig. 8 Simulation results are plotted for the vocal tract condition of D5 in terms of the temporal waveforms of the acoustic pressure at the vocal tract entrance (p_s), glottal volume flow (u_g), and the displacement of the masses (y_1 and y_2), where the blue line represents the displacement for the lower mass and the red line represents displacement for the upper mass. The natural frequency of the vocal folds was 630 Hz. Simulation was performed with (right) and without (left) the interaction. Broken lines with letters (d) and (e) show the type of phase relation among the glottal variables (see text).

in the current study using morphological measurement of the vocal tract and a computer simulation of the voice production process from a physical perspective, investigating how the acoustic load of the vocal tract affects the oscillatory behavior of the vocal folds during phonation. To reproduce a realistic situation, the vocal tract of a soprano was scanned using MRI while she produced a range of notes. We found that the participant tuned their vocal tract resonance in the proximity of the fundamental frequency when the note became higher, as reported in the literatures [12–14]. Because the source-filter interaction is caused by the acoustic pressure near the glottis and this acoustic pressure is determined by the vocal-tract input impedance, we were able to evaluate the interaction effect accurately using morphological data obtained via MRI measurement. A number of simulation studies have been performed using morphological data taken for spoken vowels [3,5], but we gathered vocal-tract data of sung vowels in three dimensions and examined the interaction effect using the observed vocal-tract configurations.

Computer simulation was then performed using acoustic models of the vocal tract [19] and vocal folds [22], where the acoustic pressure at the entrance of the vocal tract was used to incorporate the effects of the interaction on the glottal volume flow and mechanical movements of the vocal folds [5]. The natural frequency of the vocal fold model and the lung pressure were the parameters for controlling the vibration frequency of the folds, and the values of these parameters were determined so that the simulated vibration frequency agreed with the actual fundamental frequency of three musical notes produced by the participant during the MRI measurement. Finally, simulation experiments were performed for every combination of four vocal tract conditions (input impedances) and three vocal fold conditions (i.e., the values of natural frequency and lung pressure).

Using this particular simulation setup, we were able to examine the effects of the interaction for various relationships between the impedance peak of the vocal tract and the vibration frequency of the vocal folds. In addition, the change in vibration frequency was examined when the interaction was switched on and off. We found that this change was significant when the natural frequency of the vocal folds was in the proximity of an impedance peak, and the interaction generally decreased the vibration frequency. This was typical when the vocal tract condition was set based on the morphological data of A3 and the natural frequency of the vocal folds was 340 Hz, or when the vocal tract condition was D5 and the natural frequency was 630 Hz. In particular, for the D5 vocal tract condition, 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.

Finally, we examined simulated temporal waveforms of the glottal variables, including the acoustic pressure at the entrance of the vocal tract, glottal volume flow, and displacement of each vocal-fold mass. As expected, the amplitude of the acoustic pressure (and hence the strength of the interaction) was greatest when the natural frequency of the vocal folds was in the proximity of an impedance peak. For the A3-340 Hz setting mentioned above, this strong interaction decreased the vibration frequency to a high degree. Interestingly, due to this lowering of the vibration frequency, the phase relation between the acoustic pressure and the glottal flow or the mass displacement became in favor of phonation. This was also true in the D5-630 Hz setting. Furthermore, we found that the temporal patterns of the glottal flow and mass displacement fluctuated to a large degree under the influence of the source-filter interaction.

The results of our experiments were along the lines of those obtained by the forgoing simulation studies [3–6]. First, the interaction became prominent as the fundamental frequency approached an impedance peak. Second, the interaction produced unstable behavior of the vocal folds

and related glottal variables. In addition to these known properties of the interaction, the effect of the interaction was clearly demonstrated by examining the vibration frequency of the vocal folds and the phase relation among the glottal variables with and without the interaction, as mentioned above. The source-filter interaction generally decreased the vibration frequency. The mechanism of this frequency lowering can be explained as follows.

Suppose that the vibration frequency of the vocal folds was lower than an impedance peak. Under the inductive load of the vocal tract, the vocal folds began to open when the acoustic pressure at the vocal-tract entrance had a positive peak. The positive pressure could assist the vocal folds in opening widely, resulting in the lengthening of the fundamental period of the vocal fold oscillation and lowering of the vibration frequency. On the other hand, when the acoustic load of the vocal tract was capacitive in the A3-630 Hz setting, for example, the vocal folds began to close when the acoustic pressure had a positive peak. The positive pressure could then interfere with the closing movement of the folds, resulting in the lengthening of the fundamental period also.

In relation to the soprano singing, our findings suggest that the formant tuning could have both pros and cons. If the frequency of a formant is adjusted in the proximity of the fundamental frequency, the glottal sound source will be enhanced with ease by the corresponding vocal tract resonance. However, this proximity relationship may cause voice instabilities. In particular, sopranos typically use formant tuning around 600 to 1,000 Hz [12–14]. The current experiments revealed that the magnitude of the impedance peak was greater when its frequency was higher, resulting in a stronger effect of the interaction, at least for the /i/ vowel used in this study.

In addition, the interaction decreased the fundamental frequency of voice to a large degree as the natural frequency approached an impedance peak. 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 and lung pressure, particularly when the fundamental frequency is as high as the frequency of a vocal-tract resonance. Fine adjustment of the fundamental frequency may become difficult under a strong source-filter interaction, and it requires the simultaneous control of the larynx and the vocal tract. Further study is therefore needed for exploring such synergistic control of the speech organs when an experienced soprano adjusts the fundamental frequency while managing undesirable interaction effects.

ACKNOWLEDGMENT

This research was party supported by JSPS KAKENHI Grant Number JP16K00242.

REFERENCES

[1] G. Fant, *Acoustic Theory of Speech Production with Calculations Based on x-ray Studies of Russian Articulations*, 2nd printing (Mouton, The Hague, 1970).

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

[3] 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.*, **121**, 1119–1129 (2007).

[4] 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.*, **122**, 519–531 (2007).

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

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

[7] I. R. Titze and A. S. Worley, "Modeling source-filter interaction in belting and high-pitched operatic male singing," *J. Acoust. Soc. Am.*, **126**, 1530–1540 (2009).

[8] I. R. Titze, T. Riede and P. Popolo, "Nonlinear source-filter coupling in phonation: Vocal exercises," *J. Acoust. Soc. Am.*, **123**, 1902–1915 (2008).

[9] 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.*, **129**, 326–339 (2011).

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

[11] K. N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, MA, 1998), pp. 287–288.

[12] J. Sundberg, "Formant technique in a professional female singer," *Acustica*, **32**, 89–96 (1975).

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

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

[15] M. Rokkaku, K. Hashimoto, S. Imaizumi, S. Niimi and S. Kiritani, "Measurements of the three-dimensional shape of the vocal tract based on the magnetic resonance imaging technique," *Ann. Bull. RILP*, **20**, 47–54 (1986).

[16] T. Baer, J. C. Gore, L. C. Gracco and P. W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *J. Acoust. Soc. Am.*, **90**, 799–828 (1991).

[17] B. H. Story, I. R. Titze and E. A. Hoffman, "Vocal tract area function from magnetic resonance imaging," *J. Acoust. Soc. Am.*, **100**, 537–554 (1996).

[18] H. Takemoto, T. Kitamura, H. Nishimoto and K. Honda, "A method of tooth superposition on MRI data for accurate measurement of vocal tract shape and dimensions," *Acoust. Sci. & Tech.*, **25**, 468–474 (2004).

[19] M. M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. Acoust. Speech Signal Process.*, **35**, 955–967 (1987).

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

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

[22] 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," *Acta Acust. united Ac.*, **84**, 1135–1150 (1998).

[23] X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands and Y. Auregan, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," *J. Acoust. Soc. Am.*, **96**, 3416–3431 (1994).

[24] T. Kaburagi, "On the viscous-inviscid interaction of the flow passing through the glottis," *Acoust. Sci. & Tech.*, **29**, 167–175 (2008).

[25] X. Sun, "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio," *Proc. ICASSP 2002*, I-333–336 (2002).