


# Vocal tract contribution to vocal intensity: Interaction between vocal fold adduction, formant tuning, and fundamental frequency

Zhaoyan Zhang<sup>a)</sup> 

Department of Head and Neck Surgery, University of California, Los Angeles, 31-24 Rehab Center, 1000 Veteran Avenue, Los Angeles, California 90095-1794, USA

## ABSTRACT:

The goal of this study was to understand the interaction between the voice source spectral shape, formant tuning, and fundamental frequency in determining the vocal tract contribution to vocal intensity. Computational voice simulations were performed with parametric variations in both vocal fold and vocal tract configurations. The vocal tract contribution to vocal intensity was quantified as the difference in the A-weighted sound pressure level between the radiated sound pressure and the sound pressure at the glottis. The results from the simulations showed that the vocal tract contribution to vocal intensity depends strongly on the amplitude of the first vocal tract resonance. Two strategies to increase vocal tract contribution to vocal intensity were identified. The first strategy was to increase vocal fold adduction, which increases the relative prominence of the harmonics near the first vocal tract resonance. The second strategy was to bring a vocal tract resonance and a nearby harmonic closer in frequency, often known as formant tuning. In this study, increasing vocal fold adduction was the primary strategy at low fundamental frequencies, whereas formant tuning was more effective at high fundamental frequencies, particularly when formant tuning involves the strongest harmonic in the voice source spectrum. © 2025 Acoustical Society of America.

<https://doi.org/10.1121/10.0039239>

(Received 23 May 2025; revised 13 August 2025; accepted 17 August 2025; published online 9 September 2025)

[Editor: James F. Lynch]

Pages: 1904–1913

## I. INTRODUCTION

While acoustic propagation in the vocal tract is a passive process, the presence of the vocal tract improves impedance matching between the glottis and the free space outside the mouth, thus improving radiation efficiency and increasing the radiated vocal intensity. Such an increase in vocal intensity is the largest when one of the vocal tract resonances is precisely tuned to a harmonic of the voice source, often known as formant tuning in voice pedagogy (Miller and Schutte, 1990; Titze, 1992; Titze and Sundberg, 1992). This is particularly the case if the harmonic being tuned to is the strongest harmonic in the voice source spectrum, which can significantly increase the radiated vocal intensity. Formant tuning is often observed in professional female singers, particularly sopranos, in both classical singing (Sundberg, 1975; Joliveau *et al.*, 2004; Köberlein *et al.*, 2021) and traditional singing (Henrich *et al.*, 2011).

For male voices, formant tuning has been observed within and above the passaggio (Miller and Schutte, 1990; Neumann *et al.*, 2005), although there are no common rules used by all male voices. For example, Henrich *et al.* (2011) showed that significantly improved matching was less common in baritones and tenors. Sundberg *et al.* (2011, 2013) showed that for most male singer subjects, the formant

frequencies remained the same or similar as the pitch was varied, such that “coincidence between the first or second formant and a harmonic appeared rather to happen by chance.”

The inconsistent use of formant tuning at lower voices is probably because at low fundamental frequencies, the harmonics are closely spaced, so at least one harmonic will usually fall sufficiently close to the first resonance to obtain some boost in sound level. In general, the vocal tract contribution to vocal intensity is determined by how close the strongest harmonic is in frequency to its nearby vocal tract resonance, as well as the strongest harmonic’s prominence in amplitude with respect to other harmonics. The closer the strongest harmonic is to its nearby resonance, the larger it will be boosted in amplitude. The more prominent the strongest harmonic is in the voice source spectrum, the more importantly its boost in amplitude contributes to the boost in overall vocal intensity. Thus, for the lower voices where the first resonance is always close to a harmonic, the vocal tract contribution to vocal intensity would depend less on formant tuning and more on the low-frequency spectral shape, or the relative prominence between the first few harmonics, particularly those harmonics near the first vocal tract resonance, which may include harmonics up to the third, fourth, or an even higher harmonic depending on the ratio between the fundamental frequency and first vocal tract resonance.

<sup>a)</sup>Email: [zyzhang@ucla.edu](mailto:zyzhang@ucla.edu)

Such interaction between the vocal tract, voice source spectral shape, and fundamental frequency in determining the radiated sound pressure level was observed in previous studies. In a theoretical study, Titze (1992) showed that fluctuations in vocal intensity with formant tuning can exceed 5 dB at high pitches but are usually smaller than 3 dB in the male speech range, where the first vocal tract resonance is close to the fourth or even higher-order harmonics. He also showed that while the vocal intensity in general increases with the fundamental frequency, the vocal intensity for a voice source with strong high-frequency harmonics is overall much higher than that for a voice source with weak high-frequency harmonics. In a more recent computational study, Herbst and Story (2022) expanded this investigation to a large range of vocal tract configurations. They showed that formant tuning only had a marginal effect on the A-weighted sound pressure level at low pitches, at which source spectral slope was the predominant factor for sound level increase. Formant tuning became more prevalent, and source strength played an increasingly marginal role as the fo increased to the upper limits of the soprano range.

A limitation of these previous studies is that they only considered a uniformly decaying harmonic voice source spectrum with a prescribed negative spectral slope. While the high-frequency component of the voice source spectrum may be adequately characterized by a uniformly decaying spectrum, the low-frequency component may rise with frequency so that the first harmonic may not always be the strongest harmonic in the voice source spectrum, particularly at low pitches. Considering that the low-frequency harmonics often dominate the high-frequency harmonics in energy, the low-frequency spectral shape may have important implications for the vocal tract contribution to the radiated sound pressure level. The purpose of this study is to provide a more systematic understanding of the contribution to vocal intensity by the vocal tract in general, and formant tuning in particular. Using a self-oscillating three-dimensional model with inherent source-filter interaction, voice simulations were performed with parametric

variations in both the vocal fold and vocal tract configurations, which allows us to systematically investigate the interaction between fundamental frequency, source spectral shape, and formant tuning in determining the vocal tract contribution to vocal intensity.

## II. METHOD

### A. Computational model and simulation conditions

A three-dimensional body-cover, continuum model of voice production as developed in our previous studies (Zhang, 2016, 2023) was used in this study. The details of the model formulation can be found in these previous studies. A sketch of the three-dimensional vocal fold model with its geometric control parameters is shown in Fig. 1. The geometric control parameters include vocal fold length  $L$  in the anterior–posterior (AP) direction, medial-lateral widths  $D_b$  and  $D_c$  of the body and cover layers, medial surface vertical thickness  $T$ , and resting glottal angle  $\alpha$  in the horizontal plane. Note that the two medial surfaces form a straight glottal channel in the vertical plane. The vocal folds are modeled as a transversely isotropic, nearly incompressible, linear material with the plane of isotropy perpendicular to the AP direction. The control parameters for the mechanical properties of the vocal folds include the transverse stiffness  $E_t$  in the coronal plane and longitudinal shear moduli in the body and cover layers  $G_{apb}$  and  $G_{apc}$  (see more details in Zhang, 2016). Additional model controls include the subglottal pressure and vocal tract shape. The glottal flow is modeled as a one-dimensional quasi-steady glottal flow taking into consideration viscous loss up to the point of flow separation, with the flow separation point predicted by an ad hoc geometric model (Zhang, 2017). Vocal fold contact is modeled using the penalty method by applying a penalty pressure perpendicular to the vocal fold surface when the two vocal folds are in contact to prevent vocal fold penetration into each other (Zhang, 2015). Despite simplifications made in our model to improve computational efficiency, our model has been shown to qualitatively and quantitatively

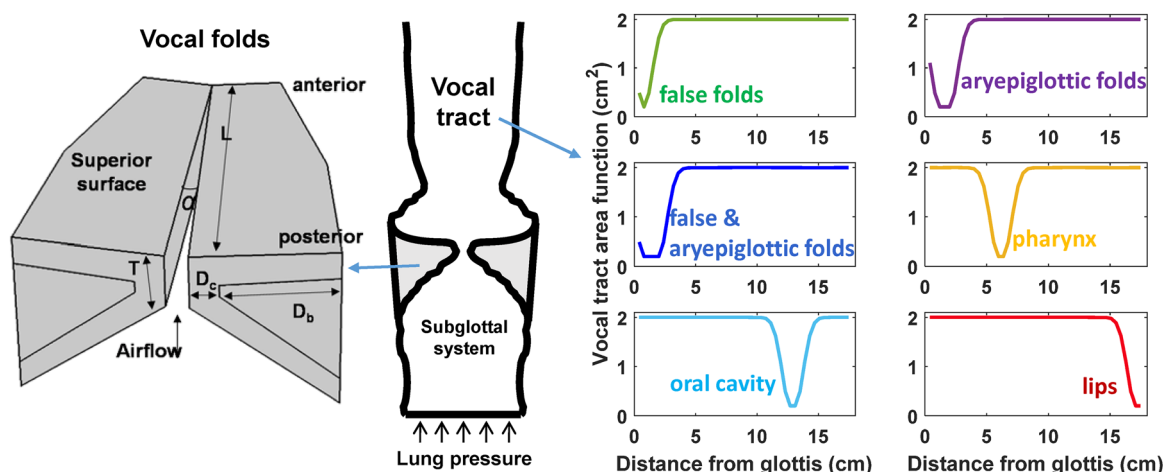


FIG. 1. A sketch of the vocal fold model and vocal tract configurations used in this study. See text and Table I for details of the model and simulation conditions.

reproduce observations from experiments and fully-resolved simulations (Zhang *et al.*, 2002; Zhang and Luu, 2012; Farahani and Zhang, 2016; Yoshinaga *et al.*, 2022; Yoshinaga and Zhang, 2025).

For the vocal tract, instead of modeling the shape for individual speech sounds, we introduced constriction of varying degrees to an otherwise uniform vocal tract (with a cross-sectional area of  $2\text{ cm}^2$ ) at five different locations where constrictions are often observed in speech production. These include the locations of the false vocal folds, aryepiglottic folds, pharynx, oral cavity, and lips. The vocal tract is modeled as a one-dimensional waveguide with a yielding vocal tract wall (Story *et al.*, 1996; Zhang, 2023). The effective mass, stiffness, and damping per unit area of the vocal tract wall were set to  $16.3\text{ kg/m}^2$ ,  $2187.0\text{ kN/m}^3$ , and  $13\,980\text{ Ns/m}^3$ , respectively (Milenkovic and Mo, 1988). The vocal tract model also includes viscous loss and kinetic pressure loss.

In this study, voice production simulations were performed with parametric variation in vocal fold geometry, stiffness, and position, subglottal pressure, and vocal tract minimal constriction area and location, as listed in Table I. These ranges are based on previous experimental and computational studies (Hollien, 1960; Hollien and Curtis, 1960; Titze and Talkin, 1979; Hirano and Kakita, 1985; Alipour-Haghighi and Titze, 1991; Story *et al.*, 1996; Wu and Zhang, 2016; Zhang *et al.*, 2017). In total 144 000 vocal conditions were simulated, with each condition simulating a 0.3-s-long sustained phonation.

## B. Data analysis

For each simulation condition, the radiated sound pressure at 30 cm from the lips was calculated from the volume flow velocity at the lips, assuming a monopole sound source mounted on an infinite baffle. The A-weighted sound pressure level (SPL)  $L$  of the radiated sound was then calculated to approximate the perceived vocal loudness and identify potential strategies to increase vocal loudness. An A-weighted sound pressure level at the glottis  $L_g$  was also calculated from the glottal volume flow velocity using the same procedure, similar to that in Titze (1992). To facilitate comparison with the radiated sound pressure level  $L$ ,  $L_g$  was calculated as the sound pressure level that would be

produced at a distance of 30 cm from an infinitely baffled monopole sound source with a source strength equal to the glottal volume flow velocity.

The fundamental frequency,  $f_0$ , was extracted from the radiated acoustic pressure. From the voice source (i.e., the time derivative of the glottal flow waveform), the differences between the amplitude of the first harmonic and amplitude of the second harmonic (H1–H2), amplitude of the fourth harmonic (H1–H4), amplitude of the harmonic nearest 2 kHz (H1–H2 k), and amplitude of the harmonic nearest 5 kHz (H1–H5 k) were calculated. A spectral balance measure  $\alpha$  was also calculated as the difference between energy above 2 kHz and energy below 2 kHz in the voice source spectrum.

For each vocal tract condition, the vocal tract transfer function (the ratio between the volume flow velocities at the outlet and inlet to the vocal tract) was calculated from the acoustic response of the vocal tract to an impulse input at the vocal tract inlet. The vocal tract resonances were identified from the peaks in the vocal tract transfer function. The first four vocal tract resonance frequencies (R1 to R4) and their amplitudes (A1 to A4) in the vocal tract transfer function were recorded. For each vocal fold/vocal tract condition, the degree of formant tuning (DFT) was evaluated as the distance in frequency between the corresponding vocal tract resonance and the nearest harmonic, which was further normalized by the fundamental frequency. Thus, the DFT ranges from 0 to 0.5, with the value 0 indicating perfect formant tuning and 0.5 meaning the least formant tuning. In this study, the degree of formant tuning was calculated for formant tuning involving the first (DFT1) and second vocal tract resonance (DFT2), as often observed in previous studies (Sundberg, 1975; Joliveau *et al.*, 2004; Köberlein *et al.*, 2021; Henrich *et al.*, 2011).

The vocal tract contribution to vocal intensity was calculated as  $\Delta L = L - L_g$ , or the difference in the A-weighted sound pressure level between the radiated sound and the sound at the voice source. A linear regression model was further developed to relate  $\Delta L$  to parameters characterizing the voice source, the vocal tract, and formant tuning. Initially, the model included the fundamental frequency  $f_0$ , the five spectral shape measures (H1–H2, H1–H4, H1–H2 k, H1–H5 k,  $\alpha$ ), unweighted amplitudes of the first four vocal tract resonances (peaks in the vocal tract transfer function; A1 to A4), and the two degrees of formant tuning DFT1 and DFT2. However, strong collinearity was observed between the different source spectral shape measures and between the amplitudes of the first four vocal tract resonances. Four measures (H1–H2 k, H1–H5 k, A3, and A4) were eventually removed from the model so that the variance inflation factors of the remaining parameters were all smaller than 2.5. For the final model, the independent variables include  $f_0$ , DFT1, DFT2, A1, A2, H1–H2, H1–H4, and  $\alpha$ . The linear regression model included both the main effects and two-way interactions of all the included independent variables. The effect size for each of the eight control parameters was quantified by the incremental  $R^2$  value

TABLE I. Model control parameters and range of simulation conditions.

Transverse Young's modulus	$E_t = [1, 2, 4]\text{ kPa}$
Cover AP shear modulus	$G_{apc} = [10, 30]\text{ kPa}$
Body AP shear modulus	$G_{apb} = [10, 30]\text{ kPa}$
Vertical thickness	$T = [1, 2, 3, 4.5]\text{ mm}$
Cover layer depth	$D_c = [1, 1.5]\text{ mm}$
Body layer depth	$D_b = [6]\text{ mm}$
Vocal fold length	$L = [10, 17]\text{ mm}$
Initial glottal angle	$\alpha = [-1.6^\circ, 0^\circ, 1.6^\circ, 4^\circ, 8^\circ]$
Subglottal pressure	$P_s = 100\text{--}2000\text{ Pa}$ (16 steps)
Vocal tract minimum constriction area	$[0.2, 0.4, 1, 2]\text{ cm}^2$
Vocal tract constriction location	6 configurations (Fig. 1)

calculated as follows (Tabachnick and Fidell, 2013). For each control parameter, the linear regression was repeated, excluding this control parameter and all interaction terms involving this control parameter. The effect size of this specific control parameter was then quantified by the difference in the  $R^2$  value between the full regression model and the regression model excluding terms involving the specific control parameter (both main effects and interaction effects).

### III. RESULTS

Figure 2 shows the ranges of the fundamental frequency and A-weighted radiated sound pressure level  $L$  generated by the simulations, with the color denoting the number of the strongest harmonic ( $NH_{\max}$ ) in the radiated sound spectrum. The fundamental frequency covers the typical range observed in speech and the lower range of singing, including some of the lower range where formant tuning was observed in previous studies. The lower range of the sound pressure level is comparable to the lower and medium range observed in humans, whereas the upper range is lower than that observed in the human subject study by Pabon and Ternström (2020). This difference in the upper SPL range is likely due to the small number of vocal tract shapes simulated in this study. In particular, none of the vocal tract shapes has a mouth opening area larger than  $2\text{ cm}^2$ . This relatively small mouth opening area likely limited the upper SPL range produced in our simulations. In general, the first harmonic is the strongest harmonic in the radiated sound spectrum ( $NH_{\max} = 1$ ) at conditions of low SPLs or high fundamental frequencies. Toward high SPL at low fundamental frequencies,  $NH_{\max}$  gradually increases, reaching a

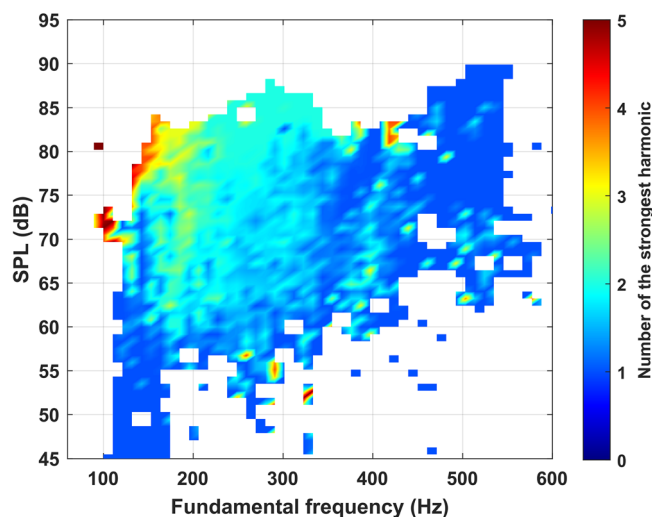


FIG. 2. Voice map of the number of the strongest harmonic ( $NH_{\max}$ ) in the radiated sound spectrum as a function of the fundamental frequency  $f_0$  and the A-weighted radiated sound pressure level (SPL). For low-intensity (below 60 dB) or high-frequency ( $f_0 > 450\text{ Hz}$ ) voices, the strongest harmonic is often the first harmonic. For low-frequency, high-intensity voices (upper left corner of the map), the first harmonic is no longer the strongest harmonic. This observation is comparable to the observation from human data as reported in Pabon and Ternström (2020).

maximum of five at the upper left corner in Fig. 2. This pattern of the  $NH_{\max}$  in the  $f_0$ -SPL space is qualitatively comparable to that reported in Pabon and Ternström (2020). Note that the region of high  $NH_{\max}$  values in our study started at around the SPL of 75 dB, which is similar to that in the study by Pabon and Ternström (2020). Our previous studies using the same model also showed that the simulated conditions are able to produce a large variety of voice qualities, including breathy, modal, pressed, vocal fry, as well as rough-sounding voices. In summary, the voice data from the simulations in this study compare reasonably with a typical vocal range observed from human subjects, except for the very high SPL range.

Figure 3 shows the vocal tract contribution to vocal intensity  $\Delta L$  as a function of the ratio between the first vocal tract resonance  $R1$  and fundamental frequency  $f_0$  for three different vocal tract configurations. For the uniform vocal tract, cyclic variations in  $\Delta L$  can be observed as the ratio  $R1/f_0$  is increased, especially at lower values of  $R1/f_0$ , with the amplitude of variation on the order of 10 dB. The peaks in  $\Delta L$  occur at integer values of  $R1/f_0$ , indicating a strong effect of resonance tuning on the radiated sound pressure level. There is also a notable peak in  $\Delta L$  at around  $R1/f_0 = 1.5$ , at which condition the  $f_0$  is related to  $R2$  at a ratio of  $R2/f_0 = 4$ , indicating resonance tuning to the second vocal tract resonance. This cyclic variation in  $\Delta L$  diminishes gradually with increasing ratio of  $R1/f_0$ . Because the  $R1$  remains constant for the same uniform vocal tract condition, this also suggests that the effect of resonance tuning on vocal intensity decreases with decreasing  $f_0$ . On the other hand, there is considerable variability at low fundamental frequencies, indicating other controls at play.

Figure 3 also shows that  $\Delta L$  varies significantly (by as large as 10 dB) across different vocal tract configurations. For a perfectly formant-tuned condition, the vocal tract with epilaryngeal narrowing provides the largest increase in sound pressure level, whereas the vocal tract with lip constriction provides the least increase in sound pressure level. This suggests that formant tuning is more effective when combined with simultaneous vocal tract adjustments that increase the overall radiation efficiency of the vocal tract, such as epilaryngeal narrowing or increased lip opening. Note that this pattern in  $\Delta L$  variation with vocal tract shape is similar to that of the amplitudes of the first vocal tract resonance  $A1$  in the vocal tract transfer function, indicating its importance in determining the vocal tract contribution to vocal intensity.

The results from the linear regression analysis of  $\Delta L$  are shown in Table II. For clarity, Table II shows all main effects but only the interaction terms with an absolute standardized coefficient larger than 0.050. The amplitude of the first vocal tract resonance  $A1$  has the largest effect, followed by  $f_0$ , DFT1,  $H1-H2$ ,  $H1-H4$ ,  $\alpha$ , DFT2, and lastly  $A2$ . In general,  $\Delta L$  increases with increasing  $A1$ , which is consistent with the observation in Fig. 3. This is expected because the larger  $A1$  is, the more strongly the harmonics around the first vocal tract resonance are amplified. There is



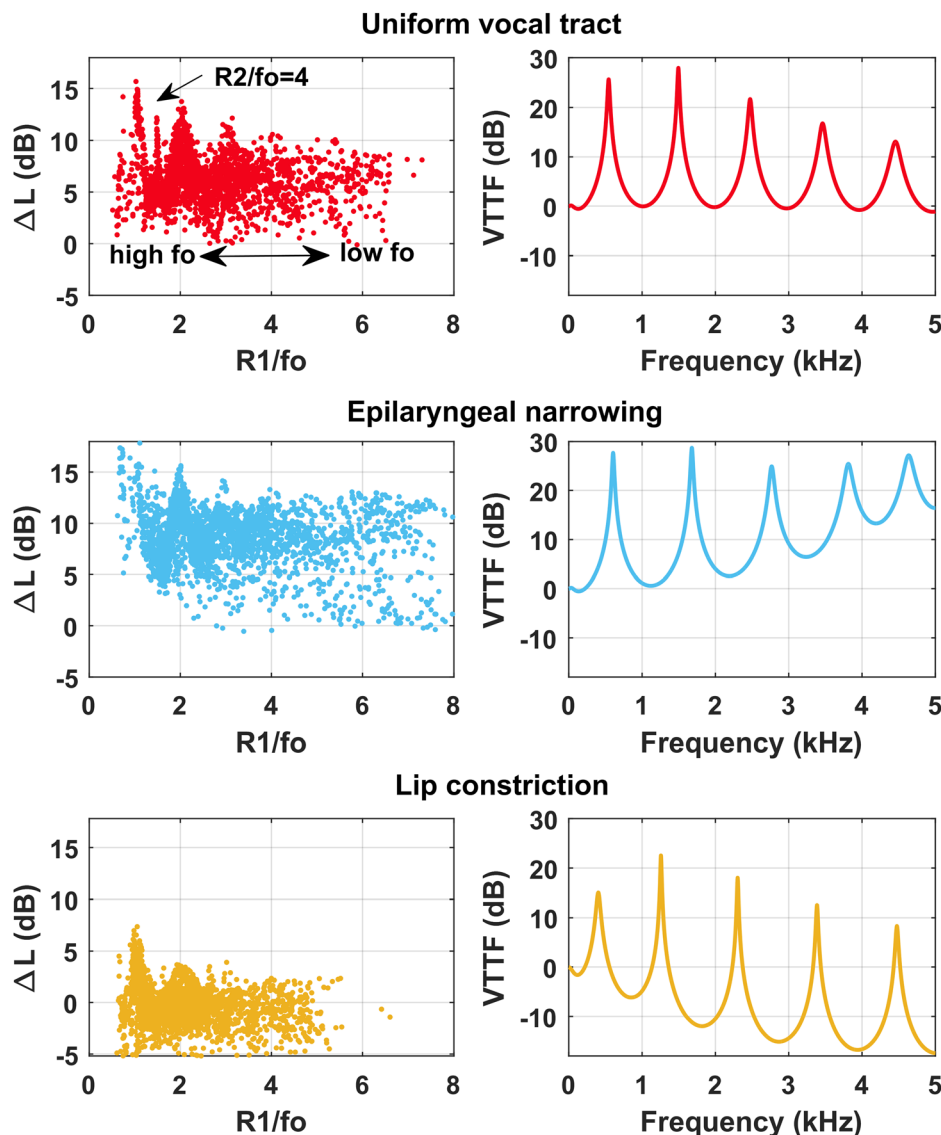


FIG. 3. Vocal tract contribution to vocal intensity  $\Delta L$  as a function of the ratio between the first vocal tract resonance and fundamental frequency ( $R1/fo$ ) for a uniform vocal tract (top), a uniform vocal tract with epilaryngeal narrowing (middle), and a uniform vocal tract with lip constriction (bottom). The right column shows the vocal tract transfer function for the corresponding vocal tract configuration. For the two vocal tracts with constrictions, the minimal vocal tract constriction area was  $0.2 \text{ cm}^2$ .

only a weak interaction between  $A1$  and  $f_0$  (with a standardized coefficient of about 0.049). Thus, epilaryngeal narrowing or increased lip opening would facilitate high-intensity voice production at both low and high fundamental frequencies.

The fundamental frequency  $f_0$  has the second largest effect, with  $\Delta L$  increasing with increasing  $f_0$ . Note that  $\Delta L$  is calculated as the difference between the radiated SPL and the SPL at the glottis, both of which were calculated taking into consideration sound radiation from the lips. Thus, it is unlikely that the effect of  $f_0$  on  $\Delta L$  is entirely due to the effect of increased radiation efficiency from the mouth with increasing frequency. The strong interaction between  $f_0$  and formant tuning (both DFT1 and DFT2) in Table II indicates that this effect of  $f_0$  is related to formant tuning. Figure 4 shows the averaged  $\Delta L$  as a function of the fundamental frequency at conditions of strong (DFT1 < 0.15) and weak (DFT1 > 0.35) formant tuning, for different values of H1–H2 and H1–H4. Figure 4 shows that the effect of  $f_0$  on  $\Delta L$  is more notable at conditions of strong formant tuning (e.g.,

DFT1 < 0.15) and a weak voice source with high H1–H2 and H1–H4 [Figs. 4(a) and 4(c)]. Under these conditions, an increase in  $f_0$  from 100 to 600 Hz can lead to an increase in  $\Delta L$  as large as 10 dB. The effect of  $f_0$  on  $\Delta L$  is much smaller with weak formant tuning or when the H1–H2 and H1–H4 are low, as often produced by increased vocal fold adduction (Zhang, 2016, 2024). In the former case, without sufficient formant tuning, increasing  $f_0$  is unable to produce a sufficient boost in  $\Delta L$  at high fundamental frequencies [note that the  $\Delta L$  does slightly increase with  $f_0$  at high fundamental frequencies in Figs. 4(b) and 4(d), but this increase is due to formant tuning to the second vocal tract resonance with DFT2 < 0.2]. In the latter case, as will be further discussed later, a lower H1–H2 or H1–H4 allows the voice source to receive a larger boost in  $\Delta L$  at low fundamental frequencies, thus reducing the overall benefit of increasing  $f_0$  on  $\Delta L$ . Note that data are missing for some conditions in Fig. 4. This is because with the small number of vocal tract conditions simulated, the range of variation in the first resonance frequency is small, which limits the degree of formant

TABLE II. Linear regression results, including the standardized coefficients, standard errors (SE) of the estimated coefficients, t-statistics, and the effect size for each of the eight control parameters (see the text for details). Only interaction terms with a standardized coefficient larger than 0.05 are listed. The  $p$  values for all terms are smaller than 0.0001.

	Coefficient	SE	tStat	Effect size
fo	0.269	0.0023	117.18	0.116
DFT1	-0.220	0.0018	-125.10	0.072
DFT2	-0.073	0.0017	-42.22	0.021
A1	0.593	0.0024	247.10	0.195
A2	0.094	0.0026	35.59	0.008
H1-H2	-0.250	0.0026	-95.36	0.038
H1-H4	-0.166	0.0025	-65.18	0.029
alpha	-0.129	0.0022	-58.03	0.028
fo:DFT1	-0.172	0.0020	-85.48	
fo:DFT2	-0.124	0.0019	-66.08	
fo:H1-H2	0.066	0.0031	21.59	
fo:H1-H4	0.090	0.0028	31.69	
fo:alpha	-0.115	0.0018	-63.68	
DFT1:H1-H2	0.073	0.0024	30.20	
DFT1:H1-H4	-0.064	0.0024	-27.03	
A1:H1-H4	-0.087	0.0032	-26.41	
A2:alpha	0.077	0.0029	27.06	
$R^2 = 0.735$				

tuning that can be achieved for some fundamental frequencies (e.g.,  $f_0 = 400$  Hz).

Table II also shows a large effect of formant tuning, particularly DFT1, in determining the  $\Delta L$ . In general, the  $\Delta L$  increases with improved formant tuning (i.e., decreasing DFT1 or DFT2). Formant tuning is more effective at high fundamental frequencies and becomes increasingly less

effective with decreasing  $f_0$ , as shown in Fig. 5. At  $f_0 = 100$  Hz,  $\Delta L$  remains almost constant across different degrees of formant tuning (Fig. 5), indicating a negligible boost in sound pressure level even with perfect formant tuning at this fundamental frequency. Two examples of the reduced effectiveness of formant tuning at low fundamental frequencies are provided in Fig. 6. Figures 6(a) and 6(b) show two voices with a relatively weak voice source with large values in H1–H4, whereas Figs. 6(c) and 6(d) show two voices with a relatively strong voice source with low values in H1–H4. In both cases, formant tuning only results in 1–3 dB increase in  $\Delta L$  despite a similarly shaped voice source. In contrast, for the voices with a higher fundamental frequency in Figs. 6(e) and 6(f), formant tuning leads to a boost of 6–9 dB in  $\Delta L$ .

Table II also shows significant effects of the low-frequency source spectral shape (H1–H2 and H1–H4) on the vocal tract contribution to vocal intensity, which increases with decreasing H1–H2 and H1–H4. This effect of H1–H2 and H1–H4 can also be observed in Fig. 6. For example, the  $\Delta L$  in Figs. 6(c) and 6(d) are slightly larger than those in Figs. 6(a) and 6(b). Strong interaction is observed between the low-frequency spectral shape, formant tuning, and fundamental frequency, as illustrated in Fig. 7. At low fundamental frequencies, decreasing H1–H2 and H1–H4, which can be achieved through increasing vocal fold adduction, is the primary means to increase  $\Delta L$ , whereas formant tuning is the least effective. In contrast, formant tuning is much more effective at high fundamental frequencies, with vocal fold adduction playing a facilitating role. When formant tuning involves the first vocal tract resonance, it is more effective when paired with reduced vocal fold adduction, which

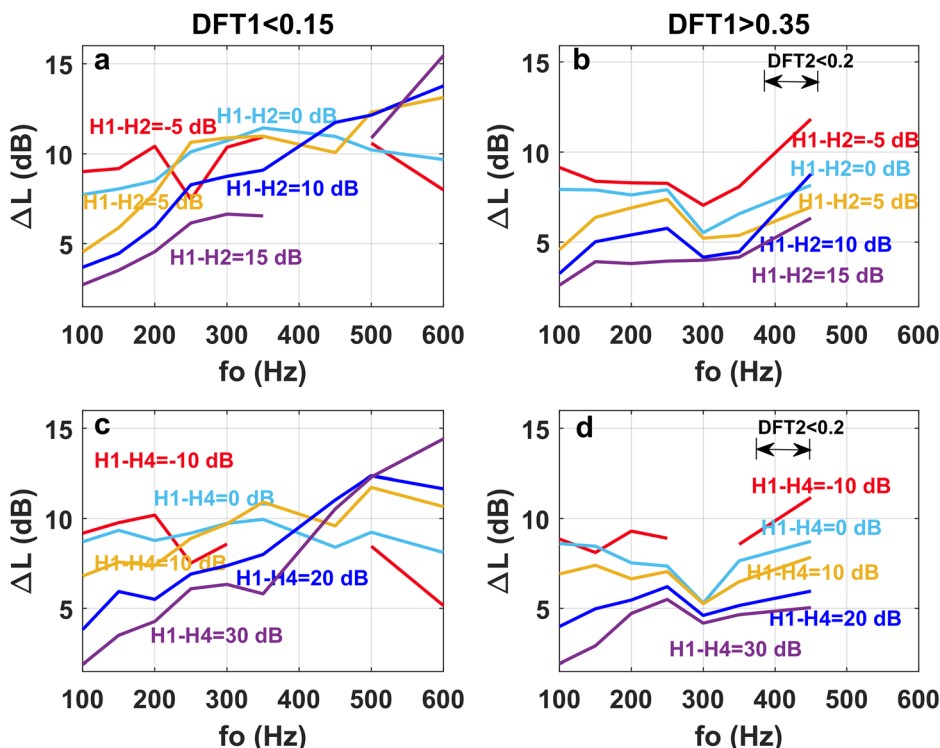


FIG. 4. Averaged values of the vocal tract contribution to vocal intensity  $\Delta L$  as a function of the fundamental frequency ( $f_0$ ) at conditions of strong (left) and weak (right) formant tuning for different values of H1–H2 and H1–H4. The effect of  $f_0$  on  $\Delta L$  is the most notable at conditions of strong formant tuning and large values of H1–H2 and H1–H4.

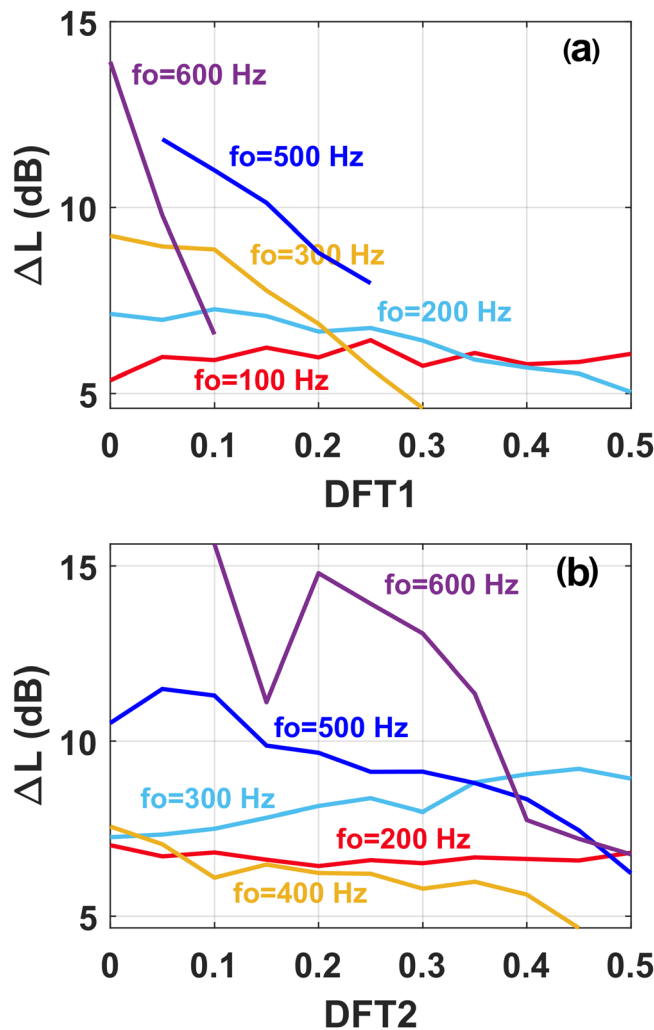


FIG. 5. Interaction between formant tuning and fundamental frequency in controlling  $\Delta L$ . Formant tuning is more effective at high fundamental frequencies.

is associated with high H1-H2 and H1-H4, as shown on the left-hand side (DFT1 < 0.25) of Figs. 7(b) and 7(d). In contrast, on the right-hand side of Figs. 7(b) and 7(d), where DFT1 > 0.3, the  $\Delta L$  appears to increase with increasing DFT1 (or reduced degree of formant tuning to the first resonance). However, this increase in  $\Delta L$  is due to formant tuning to the second vocal tract resonance, as the DFT2 decreases with increasing DFT1 and remains below 0.2 in this region. In this case, when formant tuning involves the second vocal tract resonance at high fundamental frequencies, formant tuning is more effective when paired with strong vocal fold adduction, which produces low H1-H2 and H1-H4.

The vocal tract contribution to vocal intensity also increases with decreasing  $\alpha$  (Table II), as shown in Fig. 8. This effect is likely because as the high-frequency harmonics become stronger (increasing  $\alpha$ ), this gradually reduces the prominence of the low-frequency harmonics and their contribution to the overall vocal intensity, thus decreasing the overall boost in vocal intensity. Figure 8(b) also shows that the  $\Delta L$  increases with increasing A2, as expected. There is an interaction between  $\alpha$  and A2,

with the effect of  $\alpha$  larger at lower values of A2. This effect is likely due to the correlation between A2 and the overall shape of the vocal tract transfer function in this study, as shown in Fig. 3. Specifically, the vocal tract with constriction at the lips, which has the lowest A2, also provides the least boost in high frequencies (Fig. 3). This further increases the relative dominance of the low-frequency harmonics at conditions of low A2, and thus the effect of  $\alpha$  on  $\Delta L$ . The generalization of this interaction effect needs to be investigated in the future when data from more vocal tract shapes are available.

#### IV. DISCUSSION AND CONCLUSIONS

The results of this study show that the vocal tract contribution to vocal intensity strongly depends on the amplitude of the first vocal tract resonance (A1), which has a larger effect on vocal intensity than the fundamental frequency, source spectral shape, and formant tuning. Thus, to maximize vocal intensity, a vocal tract shape that maximizes A1 should be adopted. This can be achieved by bringing the first and second vocal tract resonances closer to each other in frequency or reducing energy dissipation associated with acoustic propagation within the vocal tract. In this study, A1 slightly increases with epilaryngeal narrowing, as shown in Fig. 3. Although increased lip opening is not investigated in our study, previous studies have shown that increased lip opening and oral cavity expansion in a more realistic vocal tract shape can significantly increase A1 (Zhang, 2021). Increasing mouth opening has been observed in high-intensity voice production (e.g., Garnier *et al.*, 2022). With a reduced lip opening, the effect of formant tuning or increasing  $f_o$  on vocal intensity could be much reduced, as shown in Fig. 3.

With an adequately strong first vocal tract resonance, two strategies can be used to increase the vocal tract contribution to vocal intensity: vocal fold adduction and formant tuning, as clearly illustrated in Fig. 7. At low fundamental frequencies, the effective strategy to increase vocal tract contribution to vocal intensity is to increase vocal fold adduction, which decreases H1-H2 and H1-H4 (Zhang, 2016, 2024). This increases the relative prominence of the harmonics near the first vocal tract resonance and the overall radiated sound pressure level. This explains why vocal intensity increase in humans is often accompanied by an increased glottal resistance (Isshiki, 1964). Note that this effect on vocal intensity requires the presence of a vocal tract. Without a vocal tract, increasing vocal fold adduction (e.g., through increasing thyroarytenoid muscle activation) has only a small effect and sometimes even decreases the radiated sound pressure level (Tanaka and Tanabe, 1986; Luegmair *et al.*, 2014; Zhang, 2016).

At high fundamental frequencies, formant tuning (by changing  $f_o$  or vocal tract resonance or both) becomes the dominant strategy to increase vocal tract contribution to vocal intensity, with vocal fold adduction playing a facilitating role to position the strongest harmonic in the voice

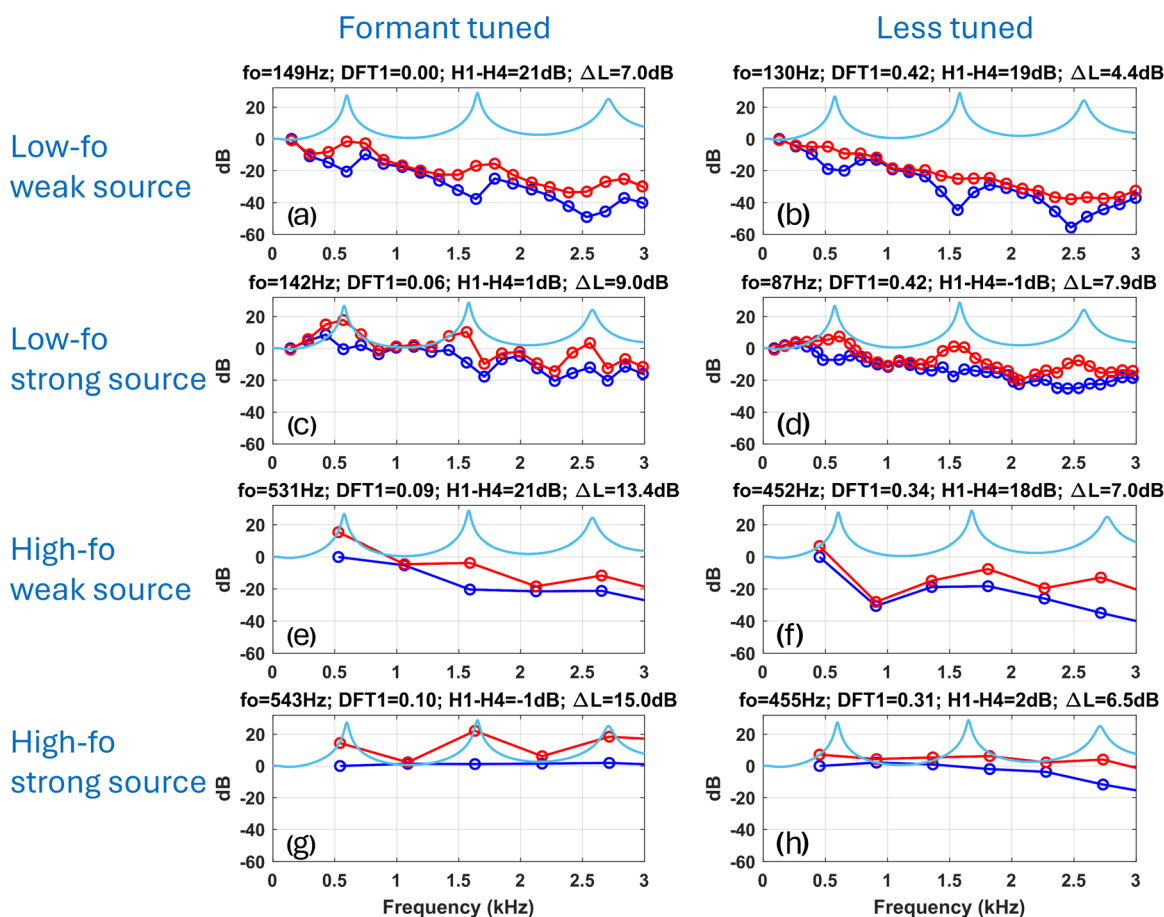


FIG. 6. Examples of the harmonic source spectrum (blue), radiated harmonic sound spectrum (red), and vocal tract transfer function (light blue) for selected conditions of fundamental frequency, formant tuning, and low-frequency source spectral shape. The harmonic spectra of both the source and radiated sound are normalized by the amplitude of the first harmonic in the source spectrum.

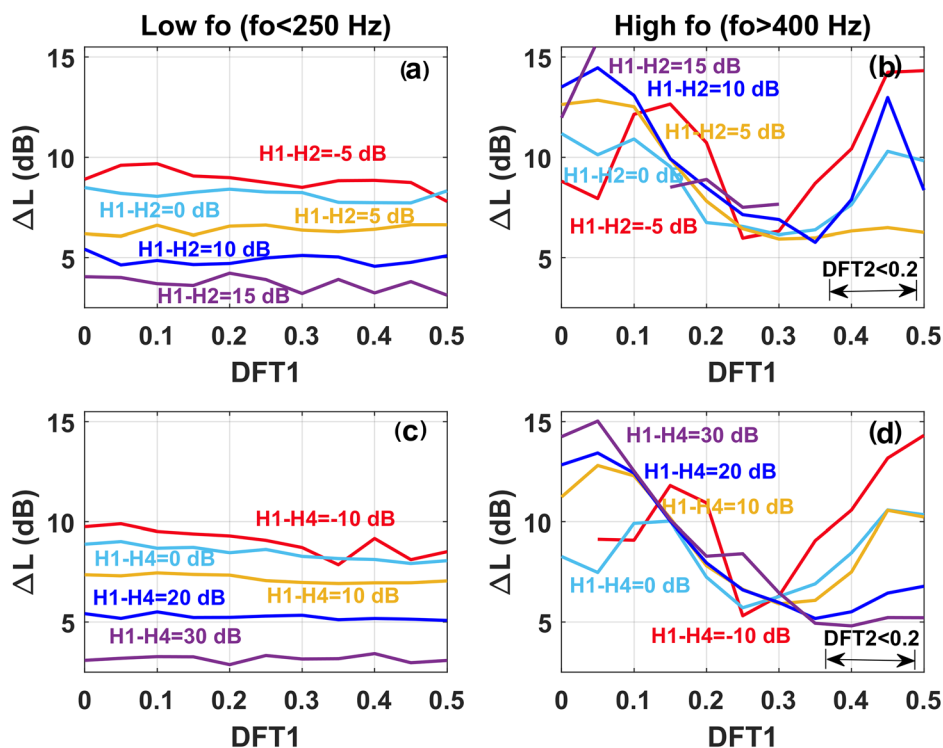


FIG. 7. Vocal fold adduction and formant tuning as two strategies to increase vocal tract contribution to vocal intensity. Averaged values of the  $\Delta L$  as a function of formant tuning (DFT1) at low (left) and high (right) fundamental frequencies for different values of H1–H2 and H1–H4.



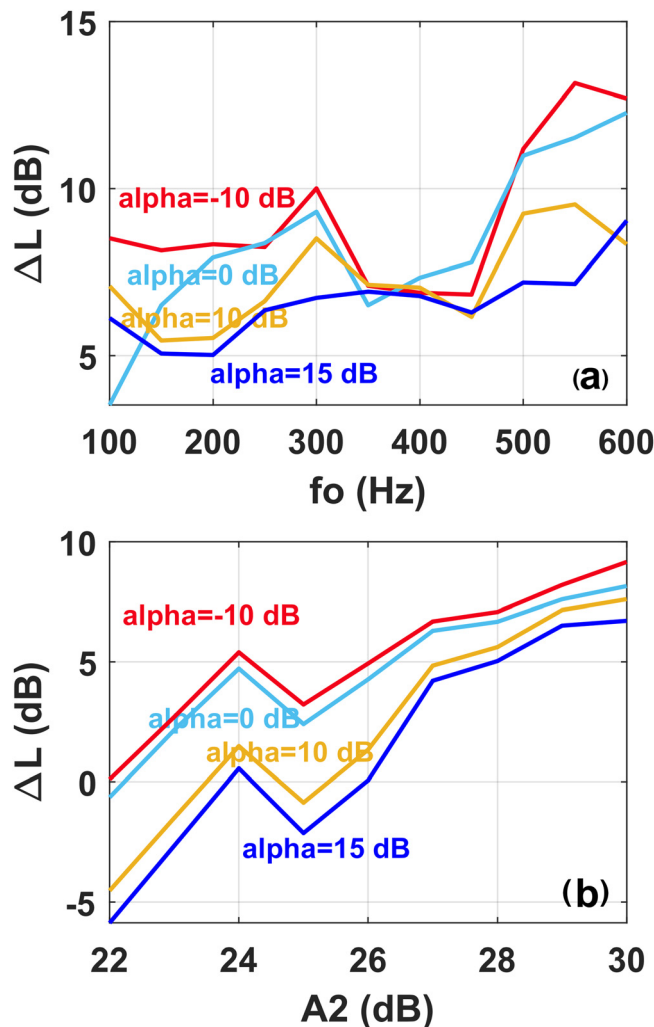


FIG. 8. The effects of alpha and  $A_2$  on the vocal tract contribution to vocal intensity, with  $\Delta L$  increasing with decreasing alpha and increasing  $A_2$ .

source spectrum closer to the vocal tract resonance involved in tuning. Specifically, when formant tuning involves the first vocal tract resonance, it is more effective when accompanied by slight reduction in vocal fold adduction, which increases H1–H2 and H1–H4 and the prominence of the first harmonic. In contrast, when formant tuning involves the second vocal tract resonance, it becomes more effective when paired with increased vocal fold adduction, which shifts the strongest harmonic to a higher harmonic closer to the second vocal tract resonance. In reality, the former case (tuning to the first vocal tract resonance) likely is more practical for voice production at high fundamental frequencies, where the vocal folds are generally thin and the produced voice source generally has a high H1–H2 and H1–H4.

A limitation of this study is the relatively small number of vocal tract configurations, which has prevented us from including the third, fourth, and even higher vocal tract resonances in the linear regression analysis. As a result, the singer's formant clustering (Sundberg, 1974), which is an important strategy to increase vocal intensity and impact voice quality, cannot be investigated in this study. While

formant tuning can be achieved by adjusting either the fundamental frequency or vocal tract resonance, adjusting the vocal tract shape to tune the first or second vocal tract resonance to a harmonic of the voice source or produce a formant clustering is more useful for singing, where the notes are often dictated by the music. The relatively small mouth opening area also limited the upper range of the radiated sound pressure level in this study. Thus, caution should be taken in extrapolating the present results to voice production at the very high intensities.

## ACKNOWLEDGMENTS

This study was supported by research Grant No. R01 DC020240 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

## AUTHOR DECLARATIONS

### Conflict of Interest

The author has no conflicts to disclose.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

- Alipour-Haghighi, F., and Titze, I. R. (1991). "Elastic models of vocal fold tissues," *J. Acoust. Soc. Am.* **90**(3), 1326–1331.
- Farahani, M., and Zhang, Z. (2016). "Experimental validation of a three dimensional reduced-order continuum model of phonation," *J. Acoust. Soc. Am.* **140**(2), EL172–EL177.
- Garnier, M., Smith, J., and Wolfe, J. (2022). "Lip hyper-articulation in loud voice: Effect on resonance-harmonic proximity," *J. Acoust. Soc. Am.* **152**(6), 3695–3705.
- Henrich, N., Smith, J., and Wolfe, J. (2011). "Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones," *J. Acoust. Soc. Am.* **129**(2), 1024–1035.
- Herbst, C. T., and Story, B. H. (2022). "Computer simulation of vocal tract resonance tuning strategies with respect to fundamental frequency and voice source spectral slope in singing," *J. Acoust. Soc. Am.* **152**(6), 3548–3561.
- Hirano, M., and Kakita, Y. (1985). "Cover-body theory of vocal fold vibration," in *Speech Science: Recent Advances*, edited by R. G. Daniloff (College-Hill, San Diego, CA), pp. 1–46.
- Hollien, H. (1960). "Vocal pitch variation related to changes in vocal fold length," *J. Speech Hear. Res.* **3**(2), 150–156.
- Hollien, H., and Curtis, F. (1960). "A laminagraphic study of vocal pitch," *J. Speech Hear. Res.* **3**(4), 361–371.
- Isshiki, N. (1964). "Regulatory mechanism of voice intensity variation," *J. Speech Hear. Res.* **7**, 17–29.
- Joliveau, E., Smith, J., and Wolfe, J. (2004). "Vocal tract resonances in singing: The soprano voice," *J. Acoust. Soc. Am.* **116**(4), 2434–2439.
- Köberlein, M., Birkholz, P., Burdumy, M., Richter, B., Burk, F., Traser, L., and Echtermach, M. (2021). "Investigation of resonance strategies of high pitch singing sopranos using dynamic three-dimensional magnetic resonance imaging," *J. Acoust. Soc. Am.* **150**(6), 4191–4202.
- Luegmair, G., Chhetri, D., and Zhang, Z. (2014). "The role of thyroarytenoid muscles in regulating glottal closure in an in vivo canine larynx model," *Proc. Mtgs. Acoust.* **22**, 060007.
- Milenkovic, P., and Mo, F. (1988). "Effect of the vocal tract yielding sidewall on inverse filter analysis of the glottal waveform," *J. Voice* **2**(4), 271–278.
- Miller, D. G., and Schutte, H. K. (1990). "Formant tuning in a professional baritone," *J. Voice* **4**(3), 231–237.

- Neumann, K., Schunda, P., Hoth, S., and Euler, H. A. (2005). "The interplay between glottis and vocal tract during the male passaggio," *Folia Phoniatr. Logop.* **57**(5–6), 308–327.
- Pabon, P., and Ternström, S. (2020). "Feature maps of the acoustic spectrum of the voice," *J. Voice* **34**(1), 161.e1–161.e26.
- Story, B. H., Titze, I. R., and Hoffman, E. A. (1996). "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.* **100**(1), 537–554.
- Sundberg, J. (1974). "Articulatory interpretation of the 'singing formant,'" *J. Acoust. Soc. Am.* **55**(4), 838–844.
- Sundberg, J. (1975). "Formant technique in a professional female singer," *Acta Acust. united Ac.* **32**(2), 89–96.
- Sundberg, J., Lã, F. M. B., and Gill, B. P. (2011). "Professional male singers' formant tuning strategies for the vowel /a/," *Logoped. Phoniatr. Vocol.* **36**(4), 156–167.
- Sundberg, J., Lã, F. M. B., and Gill, B. P. (2013). "Formant tuning strategies in professional male opera singers," *J. Voice* **27**(3), 278–288.
- Tabachnick, B. G., and Fidell, L. S. (2013). *Using Multivariate Statistics*, 6th ed. (Pearson, Boston, MA), Chap. 5.
- Tanaka, S., and Tanabe, M. (1986). "Glottal adjustment for regulating vocal intensity, an experimental study," *Acta Otolaryngol.* **102**, 315–324.
- Titze, I. R. (1992). "Acoustic interpretation of the voice range profile (phonetogram)," *J. Speech. Lang. Hear. Res.* **35**(1), 21–34.
- Titze, I. R., and Sundberg, J. (1992). "Vocal intensity in speakers and singers," *J. Acoust. Soc. Am.* **91**(5), 2936–2946.
- Titze, I. R., and Talkin, D. (1979). "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation," *J. Acoust. Soc. Am.* **66**(1), 60–74.
- Wu, L., and Zhang, Z. (2016). "A parametric vocal fold model based on magnetic resonance imaging," *J. Acoust. Soc. Am.* **140**(2), EL159–EL165.
- Yoshinaga, T., and Zhang, Z. (2025). "Effects of false vocal fold adduction and aryepiglottic sphincter narrowing on the voice source in a three-dimensional voice production model," *J. Acoust. Soc. Am.* **157**(4), 2408–2421.
- Yoshinaga, T., Zhang, Z., and Iida, A. (2022). "Comparison of one-dimensional and three-dimensional glottal flow models in left-right asymmetric vocal fold conditions," *J. Acoust. Soc. Am.* **152**(5), 2557–2569.
- Zhang, Z. (2015). "Regulation of glottal closure and airflow in a three-dimensional phonation model: Implications for vocal intensity control," *J. Acoust. Soc. Am.* **137**(2), 898–910.
- Zhang, Z. (2016). "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model," *J. Acoust. Soc. Am.* **139**(4), 1493–1507.
- Zhang, Z. (2017). "Effect of vocal fold stiffness on voice production in a three-dimensional body-cover phonation model," *J. Acoust. Soc. Am.* **142**(4), 2311–2321.
- Zhang, Z. (2021). "Vocal tract adjustments to minimize vocal fold contact pressure during phonation," *J. Acoust. Soc. Am.* **150**, 1609–1619.
- Zhang, Z. (2023). "The influence of source-filter interaction on the voice source in a three-dimensional computational model of voice production," *J. Acoust. Soc. Am.* **154**(4), 2462–2475.
- Zhang, Z. (2024). "Interaction effects in laryngeal and respiratory control of the voice source and vocal fold contact pressure," *J. Acoust. Soc. Am.* **156**(6), 4326–4335.
- Zhang, Z., and Luu, T. H. (2012). "Asymmetric vibration in a two-layer vocal fold model with left-right stiffness asymmetry: Experiment and simulation," *J. Acoust. Soc. Am.* **132**, 1626–1635.
- Zhang, Z., Mongeau, L., and Frankel, S. H. (2002). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," *J. Acoust. Soc. Am.* **112**(4), 1652–1663.
- Zhang, Z., Samajder, H., and Long, J. (2017). "Biaxial mechanical properties of human vocal fold cover under vocal fold elongation," *J. Acoust. Soc. Am.* **142**(4), EL356–EL361.