The source-filter theory of whistle-like calls in marmosets: Acoustic analysis and simulation of helium-modulated voices

Hiroki Koda
Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan

Isao T. Tokuda
Department of Mechanical Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

Masumi Wakita
Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan

Tsuyoshi Ito
Department of Human Biology and Anatomy, Graduate School of Medicine, University of the Ryukyus, Nishihara, Okinawa 903-0215, Japan

Takeshi Nishimura[a)]
Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan

(Received 11 November 2014; revised 7 May 2015; accepted 12 May 2015)

Whistle-like high-pitched "phee" calls are often used as long-distance vocal advertisements by small-bodied marmosets and tamarins in the dense forests of South America. While the source-filter theory proposes that vibration of the vocal fold is modified independently from the resonance of the supralaryngeal vocal tract (SVT) in human speech, a source-filter coupling that constrains the vibration frequency to SVT resonance effectively produces loud tonal sounds in some musical instruments. Here, a combined approach of acoustic analyses and simulation with helium-modulated voices was used to show that phee calls are produced principally with the same mechanism as in human speech. The animal keeps the fundamental frequency (f_0) close to the first formant (F_1) of the SVT, to amplify f_0. Although f_0 and F_1 are primarily independent, the degree of their tuning can be strengthened further by a flexible source-filter interaction, the variable strength of which depends upon the cross-sectional area of the laryngeal cavity. The results highlight the evolutionary antiquity and universality of the source-filter model in primates, but the study can also explore the diversification of vocal physiology, including source-filter interaction and its anatomical basis in non-human primates. © *2015 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4921607]

[ANP] Pages: 3068–3076

I. INTRODUCTION

The source-filter theory well explains the acoustic and physiological mechanisms of human speech production (Chiba and Kajiyama, 1941; Fant, 1960; Titze, 1994). The sound source is generated by vibration of the bilateral vocal folds (VFs), the acoustic properties of which are characterized by the fundamental frequency (f_0) and its higher harmonics. The supralaryngeal vocal tract (SVT) serves as a filter to amplify the harmonics of f_0 near the formants—the resonance frequencies of the SVT—and to suppress the others. The sound wave is radiated from the lips of the mouth and is partially reflected back to the glottis through the SVT (Titze, 1994, 2006). The source-filter theory of voice production has been applied successfully to human speech, in that the VF vibration is only weakly influenced by the SVT and f_0 is changeable independently from the SVT acoustics (Chiba and Kajiyama, 1941; Fant, 1960; Titze, 1994). By contrast, rigid source-filter interaction, as seen in some musical instruments (e.g., woodwinds), implies that the VF vibration is inevitably influenced by the SVT, whose resonances primarily determine f_0 (Fletcher and Rossing, 1998). Such a strong interaction (hereafter referred as source-filter coupling) prevents flexible and sophisticated modifications of the tone of the voice as seen in human speech.

The common marmoset, *Callithrix jacchus,* is a dwarfed species of New World monkeys (NWMs) inhabiting the dense tropical forests of the north-east of South America (Fleagle, 2013). Morphological and behavioural features in extant and fossil NWMs indicate that callitrichines—including marmosets and tamarins—are "phyletic dwarfs," in that a general reduction in body size appeared in this group as a derived characteristic (Plavcan and Gomez, 1993; Kay, 1994). While not well known, they live in family groups that include a dominant breeding pair and their offspring and relatives, defending their home range against rival family groups (Hubrecht, 1985; Stevenson and Rylands, 1988). Despite having such a small body, callitrichines forage in larger areas than the other NWMs (Nunn and Barton, 2000).

Common marmosets often use varied calls. One of their long-distance calls, termed a "phee" call (a loud shrill or loud phee), is observed both in wild and captive animals,

[a)]Electronic mail: nishimura.takeshi.2r@kyoto-u.ac.jp

 0001-4966/2015/137(6)/3068/9/$30.00

probably contributing to territorial advertisements against antagonistic neighbouring families and/or cohesion of their own family members (Norcross *et al.,* 1994; Bezerra and Souto, 2008; Roy *et al.,* 2011). Such a long-distance call is often found in other non-human primates such as the "song" of gibbons inhabiting the dense forest canopies of South-East Asia (Marshall and Marshall, 1976; Geissmann, 2000). The phee calls are whistle-like, with few modulations in frequency and amplitude during a single utterance (Bezerra and Souto, 2008; Roy *et al.,* 2011). f_0 of this call is stable at around 6000–8000 Hz and is greatly amplified, whereas the upper harmonics are strongly attenuated (Bezerra and Souto, 2008; Roy *et al.,* 2011). This strongly suggests that the f_0 location is close to one of the formants (F_n)—probably F_1—if they use SVT resonance to produce such whistle-like calls, as seen in human soprano singers (Sundberg, 1975). Nevertheless, such a high F_1 value corresponds to f_0 for a simple tube of around 1.0 to 1.4 cm, which is much shorter than the SVT length in adult common marmosets (around 2.5 to 3 cm). Unfortunately, the acoustical study of such high-pitched calls recorded in normal atmospheres is limited to show SVT resonance and the degree of independence between the source and filter (Koda *et al.,* 2012).

The physiological mechanisms of animal vocalization are often examined by the acoustics of voices recorded in a helium-enriched atmosphere: so-called "helium voices" (Nowicki, 1987; Rand and Dudley, 1993; Koda *et al.,* 2012; Madsen *et al.,* 2012). Under helium-enriched conditions, the sound velocity is increased. For vocalizers using SVT resonance when breathing helium, all formants of their voices inevitably shift upwards without any active control of muscle activities in SVT (Nowicki, 1987; Rand and Dudley, 1993; Koda *et al.,* 2012). As a first hypothesis, high source-filter independence should keep the same f_0 value, whereas only the formants are shifted upward in the helium-enriched atmosphere. Because the relationship between f_0 and its formants is destroyed, the intensity of f_0 should be reduced significantly (Nowicki, 1987; Koda *et al.,* 2012). In a second hypothesis, source-filter coupling should shift f_0 upward to a similar degree to the formants, preserving the relation between f_0 and formants in the helium-enriched conditions. In this case the intensities of f_0 harmonics should be maintained (Campbell and Murtagh, 1968). In fact, the acoustic analyses of helium voices demonstrated successfully that sophisticated tuning of both F_1 and f_0 produces the pure-tone-like voices of gibbons' songs, which are regulated independently (Koda *et al.,* 2012).

Here we examined the acoustics of phee calls for common marmosets in normal air and in helium-enriched atmospheres. We examined the location of f_0 and the intensities of f_0 and second harmonics ($2f_0$) from the mean power spectrum in both conditions. We also performed acoustic simulation using models of the marmoset SVT with varied topologies of the laryngeal vestibular cavity, to evaluate the degree of source-filter interactions in this call type. We discuss the physiological mechanisms and morphological contributions that produce these loud whistle-like voices in this dwarfed NWM.

II. MATERIALS AND METHODS

A. Ethics

All experiments were carried out in accordance with the third edition of the Guidelines for the Care and Use of Laboratory Primates at the Primate Research Institute of Kyoto University (KUPRI), and the experimental protocol was approved by the Animal Welfare and Care Committee of the same institute (Permit No. 2013-102).

B. Subject animals

We used three male common marmosets, *Callithrix jacchus,* born and reared at the KUPRI: Cj190, 5 years of age (yr), 0.38 kg; Cj195, 4 yr, 0.38 kg; Cj196, 4 yr, 0.41 kg. Cj190 was examined used with a paired female subject, Cj191, that stimulated the vocalizations of Cj190.

C. Apparatus and procedures

Subject vocalizations were recorded with the microphone covering frequency ranges of ultrasonic vocalizations (USVs) of 10 Hz–200 kHz (Model CM16/CMPA; SASLab Pro. software; Avisoft Bioacoustics, Berlin, Germany) in a sound-attenuated chamber. The subject was placed in a small cage (300 mm wide × 300 deep × 450 mm high), and the microphone was set ~15 cm from the cage. The microphone was connected to an audio interface for digitalization of USVs (Model UltraSoundGate 116 Hme; Avisoft Bioacoustics), and the sounds were recorded at a sampling rate of 250 kHz with 16-bit resolution. The gas concentrations of oxygen and helium, temperature, and humidity were always monitored during experiments by gas concentration meters (oxygen, XO-2200; helium, XP-3140, New Cosmos Electric Co Ltd., Osaka, Japan) and a thermo-hygrometer (Weathercom EX-501, Empex Instruments Inc., Tokyo, Japan).

The subjects have no experience with experimental training of vocalizations, and occasionally produced phee calls without any control by experimenters. After putting a subject in the chamber (Cj190 together with Cj191), we first recorded vocalizations in normal air conditions, regarded as a gas mix comprising 80% nitrogen and 20% oxygen. Then, we gradually released 3000–9000 L of a gas mix comprising 80% helium and 20% oxygen into the chamber to replace the nitrogen with helium, finally generating a heliox condition of 80% helium and 20% oxygen. The sound velocity increased from 331 m/s in the normal to 578 m/s in the final heliox condition, so the resonance frequencies of a simple tube shifted up by ~175% (Nowicki, 1987). We recorded vocalizations in varied atmospheric conditions during a session. A single session was performed once a day for any subject, and two or three sessions were conducted for each subject.

D. Acoustic analysis

The recorded sounds were analysed using PRAAT (version 5.3.52: available from Paul Boersma and David Weenick; http://www.fon.hum.uva.nl/praat/), excluding those

recordings with sound clipping and a low signal-to-noise ratio. For a single phee call, we measured the location of f_0 using autocorrelation algorithms, and generated the mean power spectrum using the "ltas" method in PRAAT. We divided the mean power spectrum into 10-Hz bins, and quantified intensities of f_0 and $2f_0$, following the same procedures as used for analyses of bird and gibbon songs. To examine the effects of helium gas, we analysed the f_0 location and the intensity differences of f_0 and $2f_0$ (hereafter, f_0-$2f_0$ intensity difference) at concentrations of 0%–80% of helium. We performed regression analyses for these values against the helium concentration as an explanatory variable for each subject.

E. Simulation analysis

A computational model was constructed to simulate vocalization of the marmosets. To generate the formant tuning, f_0 was adjusted to the F_1. Here, the VF vibration was simulated by the two-mass model, whereas the SVT was realized by the wave-reflection model. Our model took into account the mutual interaction between VF vibration and SVT acoustics using Titze's proposed formula (Titze, 2006, 2008) for human speech and singing.

Figure 1(a) shows a schematic representation of the two-mass model (Ishizaka and Flanagan, 1972; Steinecke and Herzel, 1995). The idea of this model is to divide the VF tissue into upper and lower portions of the masses m_1 and m_2, coupled by springs. Letting $x_{1\alpha}$ and $x_{2\alpha}$ be displacements of the lower and upper masses with the index denoting either left or right side ($\alpha = l, r$), the equation of motion is

$$m_1\ddot{x}_{1\alpha} + r_1\dot{x}_{1\alpha} + k_1x_{1\alpha} + \Theta(-a_1)c_1(a_1/2l) + k_c(x_{1\alpha} - x_{2\alpha}) = ld_1P_1,$$
$$m_2\ddot{x}_{2\alpha} + r_2\dot{x}_{2\alpha} + k_2x_{2\alpha} + \Theta(-a_2)c_2(a_2/2l) + k_c(x_{2\alpha} - x_{1\alpha}) = ld_2P_2.$$

Here, k_i and r_i stand for stiffness and damping of the lower and upper masses ($l = 1, 2$), respectively, whereas k_c stands for mutual coupling between the two masses. Lower and upper glottal areas are given by $a_i = a_{0i} + l\,(x_{ir} + x_{il})$, where a_{0i} represents the prephonatory area and l corresponds to the vocal fold length, and c_i describes the collision force activated during glottal closure, where the activation function is defined with $\Theta(x) = 1\ (x > 0)$, $0\ (x \geq 0)$. For simplicity, symmetrical motion between the left and right vocal folds has been assumed ($x_{1l} = x_{1r}$, $x_{2l} = x_{2r}$). Under the assumption that the flow inside the glottis obeys the Bernoulli principle below the narrowest part of the glottis, the pressure that acts

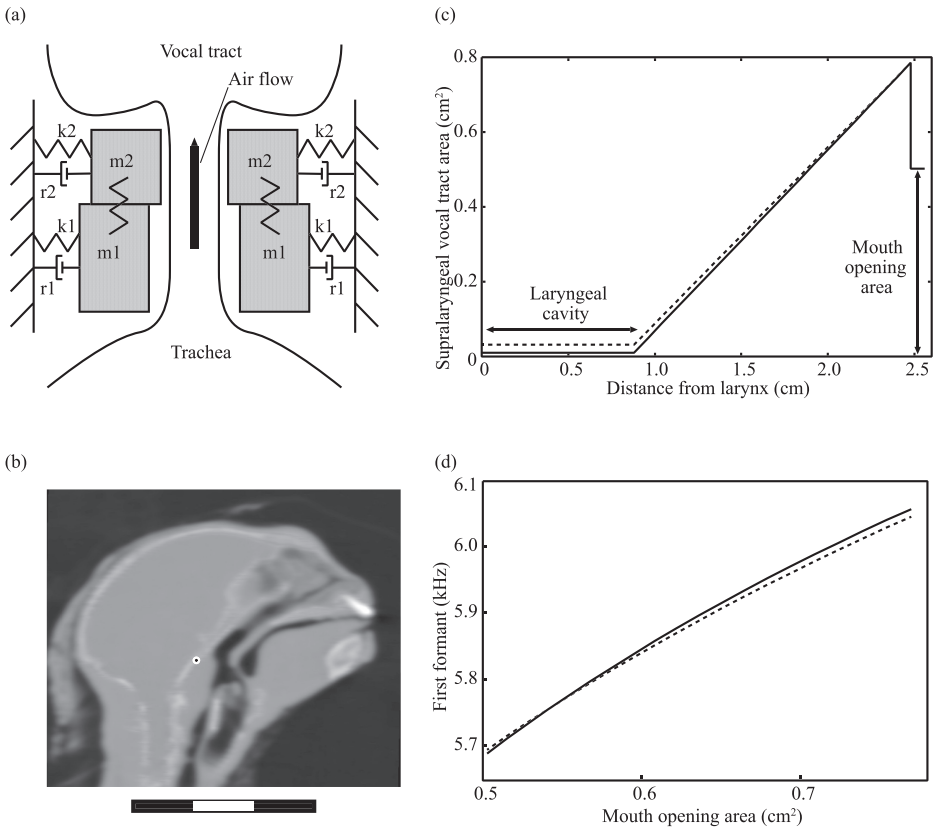

FIG. 1. Models of vocal fold vibration and SVT for acoustic simulation. (a) Schematic illustration of the two-mass model. The left and right vocal folds have a symmetrical configuration. Each vocal fold is composed of upper and lower masses coupled theoretically by linear springs. The airflow coming from the lungs is described by Bernoulli's principle below the narrowest part of the glottis. (b) Mid-sagittal computed tomography of the head for a male common marmoset, *Callithrix jacchus,* PRICT-1232. (c) The cross-sectional area functions representing whole shape of the supraglottal vocal tract with a simple uniform tube. The dotted line shows the case of strong source-filter interaction with a wide laryngeal cavity, and the solid line shows the case of weak source-filter interaction with a narrow laryngeal cavity. The open area of the mouth was used as a control parameter to change the formant frequencies. (d) Dependence of the first formant, F_1, on the opening area of the mouth, varying from 50 to 77 mm^2. The dotted line shows the case of strong source-filter interaction for supraglottal entry area (diameter 2 mm) and the solid line shows the case of weak source-filter interaction (diameter 1.1 mm).

on each mass is determined as $P_1 = P_s + (P_s - P_e)(a_1/a_{min})^2$, $P_2 = P_e$, where P_s and P_e stand for sub- and supraglottal pressures and $a_{min} = min(a_1, a_2)$.

The tension parameter Q was introduced to control the mass size and the stiffness as $m_i = m'_i/Q$, $k_i = Qk'_i$ $(i = 1, 2)$, where Q controls the f_0 value of the two-mass model linearly (Ishizaka and Flanagan, 1972). The parameter values were set as $m'_1 = 1.25$ mg, $m'_2 = 0.25$ mg, $k'_1 = 80$ kg/ms,^{2} $k'_2 = 8$ kg/ms,^{2} $k_c = 25$ kg/ms,^{2} $c_1 = 3k_1$, $c_2 = 3k_2$, $d_1 = 1$ mm, $d_2 = 0.2$ mm, $a_{01} = a_{02} = 0.2$ mm^2, $l = 3$ mm, while the damping constants were set as $r_i = 2\zeta(m_i k_i)^{1/2}$ using a damping ratio of $\zeta = 0.01$. These parameters were adjusted from the standard settings widely applied to human as well as animal vocalizations (Ishizaka and Flanagan, 1972; Amador *et al.*, 2008). The simulation results were not particularly sensitive to the parameter settings because essentially the same results were obtained within a given parameter range.

The sub- and supraglottal systems were described using the wave-reflection model (Kelly and Lochbaum, 1962; Liljencrants, 1985; Story, 1995; Titze, 2008), which is a time-domain model of the propagation of one-dimensional planar acoustic waves through a collection of uniform cylindrical tubes. The subglottal system was modelled as a simple uniform tube (diameter = 5 mm) divided into 50 cylindrical sections. The cross-sectional area function for the supraglottal tract, divided into 32 cylindrical sections, was designed to form a simple divergent shape imitating a typical vocalization of the marmoset phee call (Fig. 2). In both sub- and supraglottal systems, the section length Δz was set to 0.8 mm. The attenuation factor for the resonators was approximated as $a_k = 1 - 0.007(\pi/A_k)^{1/2}\Delta z$ (A_k: kth cylinder area). Radiation resistance and radiation inertance values at the lip were $R_r = 128\rho c/(9\pi^2 A_L)$ and $I_r = 8\rho/(3\pi^{3/2}A_L{}^{1/2})$, respectively, where the lip area A_L corresponds to the last section of the supraglottis. The value $\rho = 1.13$ mg/cm^2 represents the air density constant and $c = 0.35$ m/ms stands for the sound velocity.

To couple the sub- and supraglottal systems to the VF model, an interactive source-filter interaction was applied

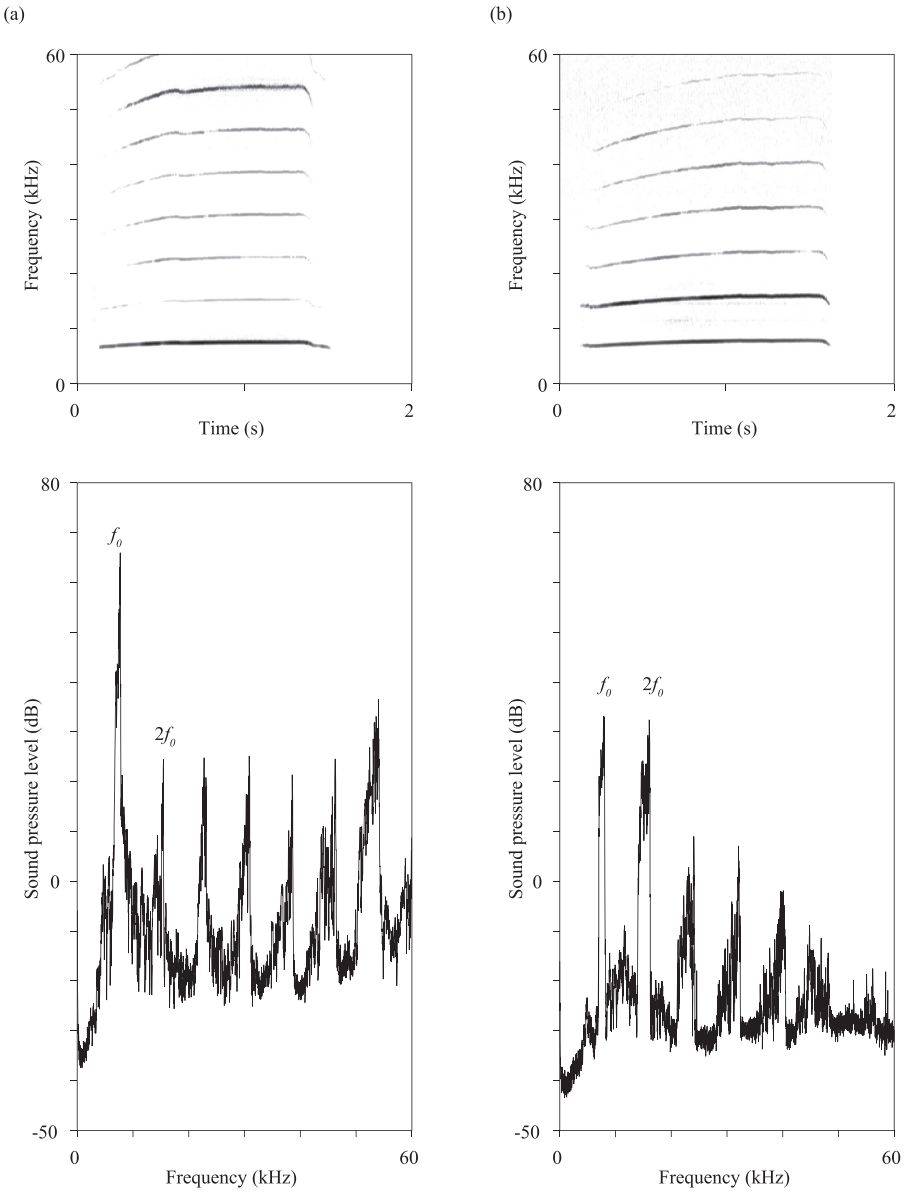

FIG. 2. Spectrogram and mean power spectrum of a sample call from subject Cj196, a male *Callithrix jacchus*. (a) Normal air condition and (b) heliox conditions.

according to Titze (2006, 2008). In this formula, the glottal flow is given by

$$u_g = \frac{a_{\min}}{k_t}\left\{-\left(\frac{a_{\min}}{A^*}\right)^2 \pm\left[\left(\frac{a_{\min}}{A^*}\right)^2+\frac{2k_t}{\rho c^2}\left(P_l+2p_s^+-2p_e^-\right)\right]^{1/2}\right\},$$

where $A^* = A_s A_e/(A_s + A_e)$, with A_s and A_e being the subglottal and supraglottal entry areas, which were set to be equal to that of the last section of the subglottal system and that of the initial section of the supraglottal system, respectively. k_t is a transglottal pressure coefficient set as 1. P_l stands for the lung pressure, whereas $p_s^{\pm}$ and $p_e^{\pm}$ represent the incident partial wave pressures in the subglottis and supraglottis (the symbol "+" denotes movement toward the mouth, whereas "−" denotes movement in the opposite direction). The sub- and supraglottal pressures are given by $P_s = P_l + p_s^+ + p_s^-$ and $P_e = p_e^+ + p_e^-$. The lung pressure was set as $P_l = 1$ kPa. To obtain the output acoustic signal, the glottal flow waveform u_g was convolved with the transmission impulse response of the supraglottal system given by the transmission line model (Sondhi and Schroeter, 1987; Story *et al.*, 2000).

The SVT topology was modelled based on computed tomographic (CT) surveys of embalmed cadavers of common marmosets [Fig. 1(b)]: the size of each segment was determined from the scans by using the OSIRIX software (Rosset *et al.*, 2004). The CT scans used here have been deposited and are available at the webpage of the Digital Morphology Museum, KUPRI (dmm.kyoto-u.ac.jp/archives/), under PRICT Nos. 1229–1232. The trachea, termed the subglottal system, was modelled as a simple uniform tube, whereas the SVT was designed to form a simple uniform tube of the laryngeal cavity with a length of 8.8 mm and a divergent shape with a length of 16.8 mm that imitated a typical marmoset vocalization [Fig. 1(c)]. The opening area of the mouth was controlled to change the formant frequencies [Fig. 1(d)]. To examine the effect of source-filter interaction, two settings were considered for the area of the laryngeal cavity, termed the supraglottal entry area (A_e). Since the laryngeal cavity directly connects the vocal fold vibration to the vocal tract acoustics, its area determines the strength of source-filter interaction (Titze, 2006, 2008). Namely, the smaller area narrows the connecting channel and thus weakens source-filter interaction. As a case of strong interaction, a diameter of 2 mm was used for the supraglottal entry area [Fig. 1(c): $A_e = \pi$ mm^2], whereas, as a case of weak interaction, a smaller diameter of 1.1 mm was used [Fig. 1(c): $A_e = 0.3025\ \pi$ mm^2]. The two settings were determined based upon the CT scan data.

Vocalization of the helium-breathing condition was simulated as follows. Under the normal air condition, the formant tuning was assumed. Namely, with respect to the first formant F_1 of the SVT, the tension parameter Q was tuned in such a way that f_0 was located close to F_1 and that the intensity difference between f_0 and $2f_0$ (hereafter, f_0-$2f_0$ intensity difference) was maximized. Next, to model the helium-enriched conditions, the sound velocity c was multiplied by 1.3 for a low helium concentration and by 1.75 for a high helium concentration (almost a final helium condition). The other parameters, including the tension parameter Q, were fixed to those tuned for the normal air condition. Insertion of helium primarily shifted the SVT acoustics, whereas the VF vibration frequency was also affected indirectly. A spectral analysis of the output signal simulated with such mistuned states gave the f_0-$2f_0$ intensity difference in the helium-enriched conditions.

III. RESULTS

A. Helium experiments

We recorded 834 phee calls from three marmosets (Cj190, 386 calls; Cj195, 125 calls; Cj196, 323 calls) in normal and helium-enriched conditions. We found that the spectral power of f_0 was amplified distinctively from the upper harmonics in normal conditions, independently of variations in f_0, in all marmosets. The mean intensity of f_0 was greater than that of $2f_0$ in normal air [Figs. 2(a) and 3; Cj190, $n = 78$, mean 22.54 dB, standard error ± 1.46 dB; Cj195, $n = 43$, 33.12 ± 0.63 dB; Cj196, $n = 54$, 39.79 ± 0.95 dB). Regardless of differences in the f_0-$2f_0$ intensity difference among the three subjects, such significant intensity differences between f_0 and $2f_0$ are greater than the theoretical prediction that the attenuations of harmonics in the laryngeal acoustics with radiation characteristics can cause a maximum difference of 12 dB between f_0 and $2f_0$ (Fant, 1960). The f_0-$2f_0$ intensity difference decreased significantly as the helium concentration increased for all marmosets [Figs. 2(b) and 3; Cj190, $F_{1,384} = 148.4$, $p < 0.001$; Cj195, $F_{1,123} = 246.9$, $p < 0.001$; Cj196, $F_{1,321} = 595.5$, $p < 0.001$].

f_0 shifted up significantly as the helium concentration increased (Cj190, $F_{1,384} = 4.87$, $p = 0.028$; Cj195, $F_{1,123} = 146.2$, $p < 0.001$; Cj196, $F_{1,321} = 84.97$, $p < 0.001$), whereas the f_0 shift was small in Cj190 compared with the other two subjects (Fig. 3). The f_0 location was 6927.02 ± 18.94 Hz for Cj190 ($n = 78$), 7635.38 ± 20.43 for Cj195 ($n = 43$), and 7648.68 ± 20.74 Hz for Cj196 ($n = 54$) in normal air (Fig. 3). The f_0 location increased on average only by 1.05 times in the heliox condition even for the two subjects Cj195 and Cj196, and such increases were much smaller than that of sound velocity, which increased by 1.75 times.

B. Mathematical simulation

The mathematical model was simulated to reproduce the increases in f_0 location and the decreases in f_0-$2f_0$ intensity differences observed in the helium-enriched conditions. The results are summarized in Fig. 4. The opening area of the mouth varied slightly from 50 to 77 mm^2, so that F_1 increased from 5600 to 6100 Hz in the normal conditions (Fig. 4). The f_0-$2f_0$ intensity difference, which was maximized by the formant tuning, was significantly reduced in the helium-enriched conditions. This is because the helium condition shifted the F_1 and thus broke the tuning between f_0 and F_1. Comparing the high with the low helium conditions, the reduction was greater in the former (Fig. 4). This is

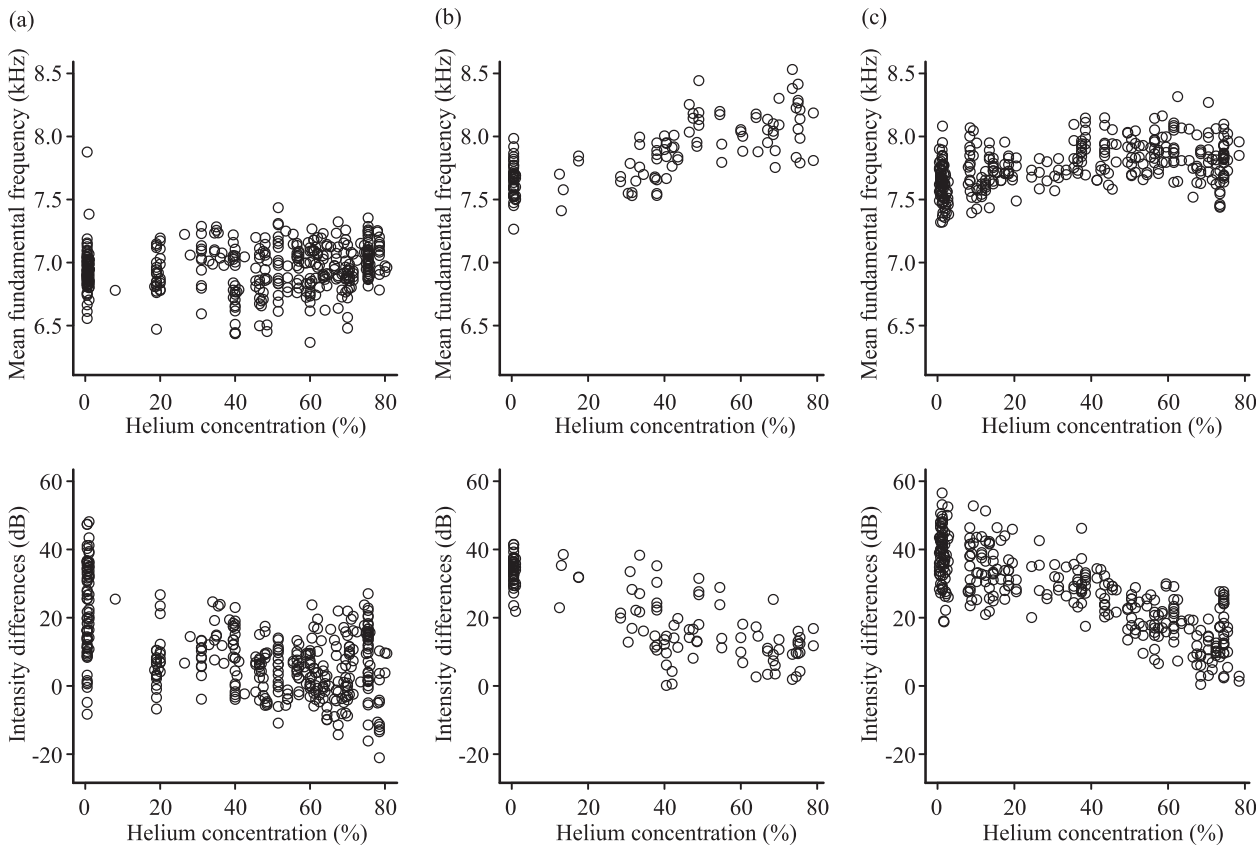

FIG. 3. Dot-plots of f_0 and f_0-$2f_0$ intensity differences plotted against helium gas concentration for subjects (a) Cj190, (b) Cj195, and (c) Cj196.

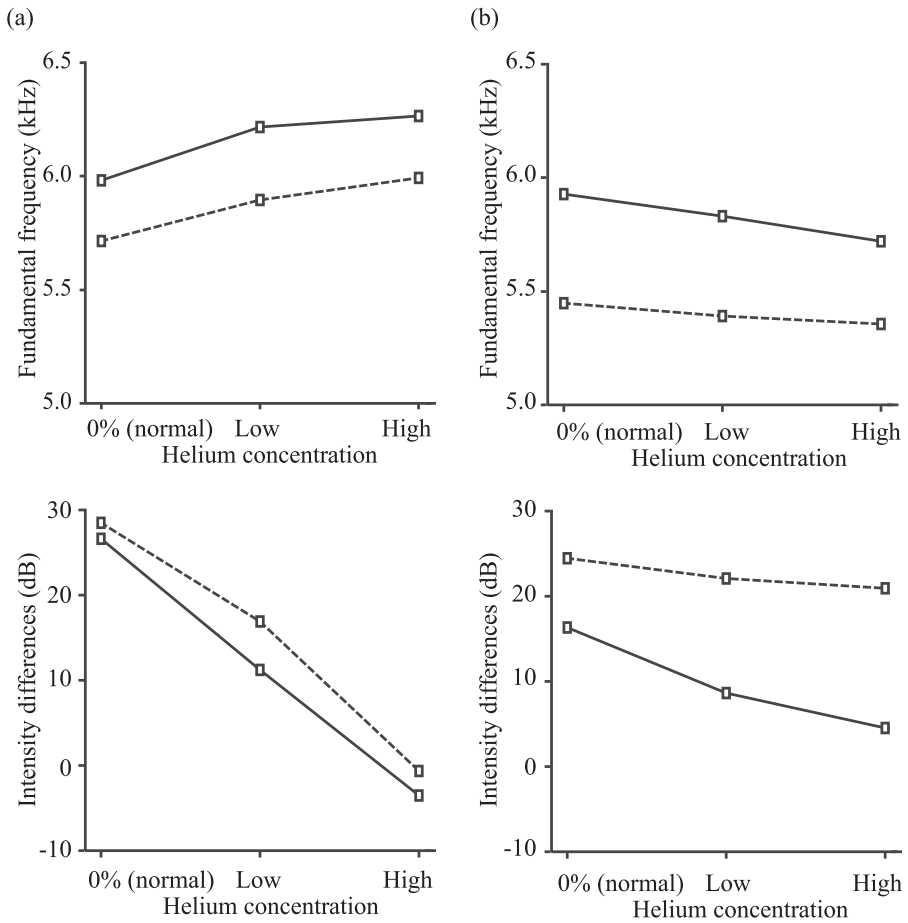

FIG. 4. Simulation of formant tuning and the effect of helium concentration. Low and high F_1 values are realized with changes in the opening area of the mouth (dotted line: low F_1, solid line: high F_1). (a) Case of strong source-filter interaction (diameter = 2 mm for supraglottal entry area); and (b) Case of weak source-filter interaction (diameter = 1.1 mm for supraglottal entry area).

because the higher helium concentration shifts F_1 to a higher frequency region, induces a larger mistuning between f_0 and F_1, and thus lowers the f_0-$2f_0$ intensity difference.

Because of source-filter interaction implemented in the model, the f_0 location also shifted from the normal to the helium-enriched conditions. The present model explains that the F_1 shift influenced f_0 as follows. Under the formant tuning assumed in the normal air condition, the negative reactance (compliance) of the SVT acoustic load, which is slightly above F_1, inhibits oscillation of the VF (Story *et al.*, 2000). As a result, f_0 was suppressed and stayed below F_1 (Adachi and Yu, 2005; Titze, 2008). Once the location of the negative reactance had been shifted to a higher frequency region by the helium, the suppressed f_0 was released and increased its value. The effect of helium on the f_0 location depended upon the strength of source-filter interaction (Fig. 4). Under strong interaction, the amount of f_0 increase was to some extent proportional to the helium concentration [Fig. 4(b)]. In the cases of low and high helium concentrations, f_0 increases were $3.5 \pm 0.4\%$ and $4.8 \pm 0.1\%$, respectively, which are in a similar range (average of 5%) observed in the experiments. These simulation results of the decreased f_0-$2f_0$ intensity difference as well as the f_0 upward shift agree quite well with the helium experiments for subjects Cj195 and Cj196 [Fig. 3(b), 3(c)]. On the other hand, under weak interaction, the f_0 location was not increased by the helium, because f_0 was nearly independent of F_1 [Fig. 4(a)]. Because helium shifted F_1 but not f_0, the tuning between F_1 and f_0 was broken so the f_0-$2f_0$ intensity difference was reduced. This well elucidates the experimental data for subject Cj190 [Fig. 3(a)].

IV. DISCUSSION

These acoustic analyses of helium voices showed that SVT resonance plays a key role in producing the phee calls in marmosets. Although f_0 shifted up slightly in the helium-enriched conditions, its shift was considerably less than that expected for formants influenced by increased helium concentrations. By contrast, the f_0-$2f_0$ intensity difference was greatly and monotonically decreased by an increase in the helium concentration. This finding indicates that marmosets normally keep F_1 close to the f_0 location to amplify f_0 exclusively. The acoustic simulation was successful in elucidating that the decreased f_0–$2f_0$ intensity difference was caused by mistuning between f_0 and F_1. Whereas f_0 was only weakly affected, F_1 was strongly shifted and separated from f_0 under helium-enriched conditions. In the sense that the magnitude of the shift was significantly different between f_0 and F_1, our study supports the view that the whistle-like phee calls are principally produced with a high degree of source-filter independence as seen in human speech, and not with a strong source-filter coupling.

Phee calls, intended as long-distance vocalizations, require effective sound transmission to attract attention from conspecific individuals widely ranging in their natural habitat: namely, dense forest with poor visibility (Bezerra and Souto, 2008; Roy *et al.*, 2011). f_0 is more powerful than any harmonic, so that its amplification with F_1 is a most reasonable solution to achieve this requirement of long-distance transmission. This physiological mechanism is basically the same as that used by the human soprano singers (Sundberg, 1975) or birds (Nowicki, 1987), and is used by gibbons to produce their loud songs (Koda *et al.*, 2012). The marmosets held the VF vibration frequency and SVT topology stable during a single utterance, keeping f_0 and F_1 tuned with each other to produce their stable phee voice. This manipulation is different from gibbons, which probably modify the VF vibration actively in co-ordination with the SVT modifications even during a single utterance (Koda *et al.*, 2012). Whereas such a simple way of manipulation could be attributed to any restrictions in neural regulation of the vocal apparatus' motions or in the cognitive ability to perceive their own audio signals in marmosets, the stable calls are a reasonable solution for a high-pitched voice. High audio frequencies are more susceptible to attenuation in dense forest (Marten *et al.*, 1977; Waser and Brown, 1986; Hauser, 1993), where the frequency modifications in high-pitched voices do not always reach the receivers correctly. Alternatively, marmosets might modify the timing of repetitions, duration of a single utterance, or pitch (f_0) to convey relevant social and cognitive information. Such a high f_0 is inevitable for this dwarfed animal (Hauser, 1993; Fitch, 1997), in contrast to medium-sized gibbons that have an f_0 value ranging from 500 to 1200 Hz. Thus, stable whistle-like calls were probably derived along with the phyletic dwarfism occurring in dense forest among common marmosets and their relatives. Even though similar ecological habitats brought about the same vocal physiology for the two phyletically distant species to produce loud and pure-tone-like voices, the difference in body size produced clear distinctions in vocal structure: stable phee calls in marmosets and melodious songs in gibbons.

It should be noted that the helium experiments demonstrated that the f_0 location was also significantly shifted upward in the helium-enriched condition in two of three subjects studied. If f_0 and F_1 were completely independent, the shift in F_1 would not have altered f_0 location. This slight increase of f_0, which was 5% on average, is within the range of f_0 fluctuation in normal air, indicating that this animal can also make such a small increase by modifying source characteristics. Given that marmosets have been shown to be sensitive to modified feedback of their own voices, there is a possibility that this 5% upward shift just reflects an arousal state increased by the change in their voices in the helium-enriched conditions. Further, despite such a slight increase, this shift might be actively made by their vocal motor control. We cannot exclude these hypotheses. The present simulation, however, implied an alternative possibility that the slight shift observed in f_0 was due to source-filter interaction, the strength of which depends upon the area of the laryngeal cavity within the SVT. This interaction of the source and filter might tightly connect f_0 and F_1 and thus strengthen the effect of the formant tuning. This mode of source-filter interaction is often involved in some forms of human vocalizations, consistent with the source-filter theory—such as high-pitched speech or singing—whereas it differs from the strong mode of source-filter coupling as seen in some musical instruments (Titze, 2006, 2008). Our analyses further

imply that the strength of source-filter interaction can be variable among individuals and among the phonation conditions, for a given single species of non-human primates. Although further empirical evidence from more subjects is needed to establish this hypothesis, such a physiological difference could in part contribute to inter-individual and inter-specific variations in the effectiveness of this long-distance vocal advertising in marmosets and their relatives. This study has provided a combined acoustic analysis and simulation of helium voices to evaluate the degree of source-filter interaction. It is known that the anatomy of the laryngeal region influences the strength of source-filter interaction (Titze, 2006, 2008). The present simulations elucidated well that anatomical modifications in the laryngeal region contribute to the varying degrees of source-filter interactions in common marmosets. The laryngeal region is probably static in topology during vocalizations in non-hominoid anthropoids including marmosets, while it is changeable and might be dynamic in hominoids. In fact, whereas the laryngeal skeleton is tightly linked to the hyoid bone in the former primates (Nishimura, 2003; Nishimura *et al.,* 2003, 2006; Nishimura *et al.,* 2008), the elements are loosely interlinked by flexible ligaments and membranes in the latter group (Nishimura, 2003; Nishimura *et al.,* 2008). Such anatomical restrictions do not allow for highly independent movements of each component and thus for topological modifications of the laryngeal cavity (Nishimura, 2003), whereas the pharyngeal configuration is rather flexibly modified for varying vocalizations in several mammals including marmosets (Fitch, 2000; Fitch and Reby, 2001; Riede *et al.,* 2005). This suggests that the degree of source-filter interaction tends to be fixed in non-human anthropoids and dynamic in hominoids including humans. Thus, the present study does not just emphasize the evolutionary antiquity and universality of the source-filter theory in non-human primates, but it also provides an approach that allows us to explore the diversifications of vocal physiology among non-human primates—including source-filter interaction and its anatomical basis. Such approach is expected to provide a new insight into the evolution of human speech physiology and anatomy.

ACKNOWLEDGMENTS

We greatly appreciate Katsuki Nakamura, Akihiro Izumi, Takumi Kunieda, and Akemi Kato for their help with animal experiments, and the staff of the Cognitive Neuroscience section and the Center of Human Evolution Model Researches of KUPRI for daily care of the animals used here. We also thank W. Tecumseh Fitch and Nobuo Masataka for many valuable comments on this study. This study was supported in part by JSPS Grants-in-Aids for Scientific Research (Grant No. 24687030 to T.N., Grant No. 22330200 to H.K., Grant Nos. 23360047, 25540074 to I.T.T.), by a JSPS Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation (to KUPRI, T.N.), and by the SPIRITS program from Kyoto University (to T.N.).

Adachi, S., and Yu, J. (**2005**). "Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing," J. Acoust. Soc. Am. **117**, 3213–3224.

Amador, A., Goller, F., and Mindlin, G. B. (**2008**). "Frequency modulation during song in a suboscine does not require vocal muscles," J. Neurophysiol. **99**, 2383–2389.

Bezerra, B. M., and Souto, A. (**2008**). "Structure and usage of the vocal repertoire of *Callithrix jacchus*," Int. J. Primatol. **29**, 671–701.

Campbell, C. J., and Murtagh, J. A. (**1968**). "Responses of various double reeds to changes in activating gas," Ann. N. Y. Acad. Sci. **155**, 351–367.

Chiba, T., and Kajiyama, M. (**1941**). *The Vowel: Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo).

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213–1222.

Fitch, W. T. (**2000**). "The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals," Phonetica **57**, 205–218.

Fitch, W. T., and Reby, D. (**2001**). "The descended larynx is not uniquely human," Proc. R. Soc. London B Biol. Sci. **268**, 1669–1675.

Fleagle, J. G. (**2013**). *Primate Adaptation and Evolution*, 3rd ed. (Academic, Amsterdam).

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments* (Springer, New York).

Geissmann, T. (**2000**). "Gibbon songs and human music in an evolutionary perspective," in *The Origins of Music*, edited by N. L. Wallin, B. Merker, and S. Brown (MIT Press, Cambridge, MA), pp. 103–123.

Hauser, M. D. (**1993**). "The evolution of nonhuman primate vocalizations: Effects of phylogeny, body weight, and social context," Am. Nat. **142**, 528–542.

Hubrecht, R. C. (**1985**). "Home-range size and use and territorial behavior in the common marmoset, *Callithrix jacchus jacchus,* at the Tapacura Field Station, Recife, Brazil," Int. J. Primatol. **6**, 533–550.

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

Kay, R. F. (**1994**). " 'Giant' tamarin from the Miocene of Colombia," Am. J. Phys. Anthropol. **95**, 333–353.

Kelly, L., and Lochbaum, C. C. (**1962**). "Speech synthesis," in *Proceedings of the Fourth International Congress on Acoustics*, Paper G42, pp. 1–4.

Koda, H., Nishimura, T., Tokuda, I. T., Oyakawa, C., Nihonmatsu, T., and Masataka, N. (**2012**). "Soprano singing in gibbons," Am. J. Phys. Anthropol. **149**, 347–355.

Liljencrants, J. (**1985**). "Speech synthesis with a reflection-type line analog," Ph.D. dissertation, Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm.

Madsen, P. T., Jensen, F. H., Carder, D., and Ridgway, S. (**2012**). "Dolphin whistles: A functional misnomer revealed by heliox breathing," Biol. Lett. **8**, 211–213.

Marshall, J. T., Jr., and Marshall, E. R. (**1976**). "Gibbons and thier territorial songs," Science **193**, 235–237.

Marten, K., Quine, D., and Marler, P. (**1977**). "Sound transmission and its significance for animal vocalization. II. Tropical forest habitats," Behav. Ecol. Sociobiol. **2**, 291–302.

Nishimura, T. (**2003**). "Comparative morphology of the hyo-laryngeal complex in anthropoids: Two steps in the evolution of the descent of the larynx," Primates **44**, 41–49.

Nishimura, T., Mikami, A., Suzuki, J., and Matsuzawa, T. (**2003**). "Descent of the larynx in chimpanzee infants," Proc. Natl. Acad. Sci. U.S.A. **100**, 6930–6933.

Nishimura, T., Mikami, A., Suzuki, J., and Matsuzawa, T. (**2006**). "Descent of the hyoid in chimpanzees: Evolution of face flattening and speech," J. Hum. Evol. **51**, 244–254.

Nishimura, T., Oishi, T., Suzuki, J., Matsuda, K., and Takahashi, T. (**2008**). "Development of the supralaryngeal vocal tract in Japanese macaques: Implications for the evolution of the descent of the larynx," Am. J. Phys. Anthropol. **135**, 182–194.

Norcross, J. L., Newman, J. D., and Fitch, W. (**1994**). "Responses to natural and synthetic phee calls by common marmosets (*Callithrix jacchus*)," Am. J. Primatol. **33**, 15–29.

Nowicki, S. (**1987**). "Vocal tract resonances in oscine bird sound production: Evidence from birdsongs in a helium atmosphere," Nature **325**, 53–55.

Nunn, C. L., and Barton, R. A. (**2000**). "Allometric slopes and independent contrasts: A comparative test of kleiber's law in primate ranging patterns," Am. Nat. **156**, 519–533.

Plavcan, J. M., and Gomez, A. M. (**1993**). "Dental scaling in the callitrichinae," Int. J. Primatol. **14**, 177–192.
Rand, A. S., and Dudley, R. (**1993**). "Frogs in helium—The anuran vocal sac is not a cavity resonator," Physiol. Zool. **66**, 793–806.
Riede, T., Bronson, E., Hatzikirou, H., and Zuberbuhler, K. (**2005**). "Vocal production mechanisms in a non-human primate: Morphological data and a model," J. Hum. Evol. **48**, 85–96.
Rosset, A., Spadola, L., and Ratib, O. (**2004**). "OsiriX: An open-source software for navigating in multidimensional DICOM images," J. Digit. Imaging **17**, 205–216.
Roy, S., Miller, C. T., Gottsch, D., and Wang, X. Q. (**2011**). "Vocal control by the common marmoset in the presence of interfering noise," J. Exp. Biol. **214**, 3619–3629.
Sondhi, M. M., and Schroeter, J. (**1987**). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE T. Acoust, Speech **35**, 955–967.
Steinecke, I., and Herzel, H. (**1995**). "Bifurcations in an asymmetric vocal-fold model," J. Acoust. Soc. Am. **97**, 1874–1884.
Stevenson, M. F., and Rylands, A. B. (**1988**). "The marmosets, genus *Callithrix*," in *Ecology and Behavior of Neotropical Primates*, edited by R. A. Mittermeier, A. F. Coimbra-Filho, A. B. Rylands, and G. A. B. da Fonseca (World Wildlife Fund, Washington, DC), pp. 131–222.
Story, B. H. (**1995**). "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph.D. dissertation, Department of Speech Pathology and Audiology, University of Iowa, Iowa City.
Story, B. H., Laukkanen, A. M., and Titze, I. R. (**2000**). "Acoustic impedance of an artificially lengthened and constricted vocal tract," J. Voice **14**, 455–469.
Sundberg, J. (**1975**). "Formant technique in a professional female Singer," Acustica **32**, 89–96.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ).
Titze, I. R. (**2006**). *The Myoelastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Iowa City).
Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.
Waser, P. M., and Brown, C. H. (**1986**). "Habitat acoustics and primate communication," Am. J. Primatol. **10**, 135–154.