Mammalian laryngseal air sacs add variability to the vocal tract impedance: Physical and computational modeling

Tobias Riede[a)]
Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi-shi, Ishikawa 923-1292, Japan; National Center for Voice and Speech, 1101 13th Street, Denver, Colorado 80204; and Department of Biology, University of Colorado at Denver and Health Sciences, Denver, Colorado 80204

Isao T. Tokuda[b)]
Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi-shi, Ishikawa 923-1292, Japan

Jacob B. Munger and Scott L. Thomson
Department of Mechanical Engineering, Brigham Young University, Provo, Utah 84602

(Received 31 October 2007; revised 17 March 2008; accepted 15 April 2008)

Cavities branching off the main vocal tract are ubiquitous in nonhumans. Mammalian air sacs exist in human relatives, including all four great apes, but only a substantially reduced version exists in humans. The present paper focuses on acoustical functions of the air sacs. The hypotheses are investigated on whether the air sacs affect amplitude of utterances and/or position of formants. A multilayer synthetic model of the vocal folds coupled with a vocal tract model was utilized. As an air sac model, four configurations were considered: open and closed uniform tube-like side branches, a rigid cavity, and an inflatable cavity. Results suggest that some air sac configurations can enhance the sound level. Furthermore, an air sac model introduces one or more additional resonance frequencies, shifting formants of the main vocal tract to some extent but not as strongly as previously suggested. In addition, dynamic range of vocalization can be extended by the air sacs. A new finding is also an increased variability of the vocal tract impedance, leading to strong nonlinear source-filter interaction effects. The experiments demonstrated that air-sac-like structures can destabilize the sound source. The results were validated by a transmission line computational model. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2924125]

PACS number(s): 43.80.Ka, 43.70.Aj, 43.70.Bk, 43.25.Ba [JAS] Pages: 634–647

I. INTRODUCTION

Study on vocal tract shape and vocal fold anatomy is indispensable for understanding the acoustic property of sounds used in vocal communications. Many aspects of nonhuman anatomy are, however, poorly understood in terms of their acoustic functions. Cavities branching off or extending from the supraglottal vocal tract are ubiquitous in mammalian species. Some of those are inflatable, for example, huge *air sac* in siamang (*Symphalangus syndactylus*) (Hill and Booth, 1957; Starck and Schneider, 1960; Schneider, 1964) or in reindeer (*Rangifer tarandus*) (Frey *et al.*, 2007). Some cavities are non-inflatable because they are surrounded partially or completely by the cartilaginous or bony material, for instance, the *Bulla hyoidea* in howler monkeys (*Alouatta spec.*) (Starck and Schneider, 1960) or the human nasal cavity. Apart from the nasal cavity, only relatively small lateral laryngeal ventricles exist in humans. It is possible that major air sacs were lost during human evolution since they exist in all four great apes, the closest relatives to the human. Only one similar voluminous anatomical structure is sometimes found in humans, the *laryngocele*, which is an extension of the lateral laryngeal ventricles. It is, however, considered as pathology.

The physiological functions of mammalian air sacs or bulla-like extensions and the reason for their absence in humans are a subject of much speculation. Hereafter, for simplicity, we use the term "air sac" to refer to both inflatable air sac and non-inflatable bulla-like extension, unless stated otherwise. We investigate possible acoustic functions of air sacs on the basis of physical model experiments as well as computational simulations. Two hypotheses on acoustic roles of the air sacs are examined in the present study. The first hypothesis is that air sacs help to optimize sound radiation by improving sound transfer through the vocal tract. This will be tested by comparing the sound emissions from physical vocal tract models with and without an air sac. The second hypothesis states that air sacs affect the position of formants. This will be tested by comparing the spectral characteristics of the sound output from the vocal tract models with and without an air sac.

The mammalian voice apparatus is mainly composed of vocal folds, which work as a sound source generator, and a vocal tract, which works as an acoustic filter. The vocal tract consists of oral and pharyngeal cavities, parts of a laryngeal cavity, and sometimes a nasal cavity (Fant, 1960). An important feature of the vocal folds is their multilayered structure (Hirano and Kakita, 1985). Mammalian vocal folds generally

[a)]Electronic mail: triede@ncvs2.org
[b)]Electronic mail: isao@jaist.ac.jp

 0001-4966/2008/124(1)/634/14/$23.00

TABLE I. Overview of four different anatomical designs of air sacs and bulla reviewed by Bartels (1905), Starck and Schneider (1960), and Hayama (1970). Further references on species-specific anatomy can be found in those three sources. A fifth design with lateral air sacs extending from the ventricles [like in (A)] with the addition of a cranial medial air sac [like in (B)] is described in detail in Hayama (1970).

(A) *Saccus laryngeus lateralis*: Each of two air sacs extends from the lateral laryngeal ventricles (*Ventriculus larynges*). It exists, for instance, in *Pongo, Gorilla*, *Pan troglotydes*, and *Cebus albifrons*. In *Symphalangus syndactulus*, the two lateral laryngeal ventricles merge ventrally from the larynx to a single expandable cavity. *Alouatta spec* has lateral air sacs in addition to a median *Bulla hyoidea*.	(B) *Saccus laryngeus medianus superior* (upper or cranial medial air sac or subhyoid sac): This is connected to the laryngeal cavity via a small short tube (*Ductus pneumaticus*). The tube enters the laryngeal cavity between the upper (rostral) edge of the thyroid cartilage and the caudoventral edge of the epiglottis. The cavity is located ventrally in a singular structure. It exists, for instance, in *Lagothrix*, and *Cercopithecus*. In Alouatta, the extension is embedded into a bulla-like enlarged hyoid bone making the *Bulla hyoidea*.
(C) *Saccus laryngeus medianus* inferior (lower or caudal medial air sac, also *Saccus intercartilagineus anterior*): This opens into the laryngeal cavity at or below the level of the vocal folds and exists, for instance, in *Leontocebus rosalia*, *Callithrix jacchus*	(D) *Saccus laryngotrachealis posterior* (dorsal air sac): This extends from the laryngeal cavity between the caudal edge of the cricoid cartilage and the first tracheal ring. It exists, for instance, in *Indri indri* and *Lemur variegata*.

consist of at least two layers: a relatively loose superficial "cover" layer and a muscular "body" layer (Hirano and Kakita, 1985). The human vocal folds also have a ligament between the cover and body layers. In some species, this ligament is not very prominent (Kurita *et al.*, 1983; Riede and Titze, 2008), although the situation is unknown for most species. Functionally, most of the vocal fold vibrations occur in the cover layer because of its high degree of compliance. The vibrations are regarded as a result of the energy transfer from the glottal airflow to the vocal fold motion (Titze, 1988). To study laryngeal aerodynamics as well as vocal fold vibrations, physical models have been developed that synthesize the flow-induced vibrations of the human vocal folds (Titze *et al.*, 1995; Chan *et al.*, 1997; Chan and Titze, 2006; Thomson *et al.*, 2005; Zhang *et al.*, 2006a; 2006b; Neubauer *et al.*, 2007). Physical models have also been constructed to simulate the vocal tract to study open questions regarding vocal tract acoustics, such as nonlinear interaction between the vocal tract and the sound source (Chan and Titze, 2006; Zhang *et al.*, 2006b), acoustic role of the piriform fossa in humans (Dang and Honda, 1997), and acoustic effect of different degrees of beak opening in birds (Fletcher and Tarnopolsky, 1999). In the present study, a modified version of the physical model of the vocal folds, which has been repeatedly examined by preceding studies (e.g., Thomson *et al.*, 2005; Zhang *et al.*, 2006a; 2006b; Neubauer *et al.*, 2007), is attached to a physical model of the vocal tract that includes the air sac. Both inflatable and non-inflatable configurations of the air sac cavity are considered. We test the hypotheses that either of these structures can help amplify sound or affect vocal tract resonances. In addition to the physical model of the vocal folds, we exploit a sweep tone generator as the sound source to determine the acoustical transfer function of the vocal tract model.

The purpose of this study is not only to determine the transfer function of the vocal tract but also to consider source-filter interactions. The sound source in the larynx and the acoustic filter in the vocal tract can interact either linearly or nonlinearly (Titze, 2008). In the case of linear coupling, the vocal source signal is produced independently of the acoustic pressures in the vocal tract. Laryngeal airflow is formed by transglottal pressure and flow pulses that mirror the time-varying glottal area. In contrast, in the case of nonlinear coupling, acoustic pressure in the vocal tract contributes to the vocal source signal production. In this case, sound radiation is affected not only by vocal tract acoustics but also by the sound source at the glottis. It is likely that the nonlinear source-filter interaction is important since specific exercises in humans have revealed that high-amplitude vocalizations are more often associated with signs of severe interaction (Titze *et al.*, 2008). The air sacs can affect the sound radiation in two ways: one is to modify the acoustical property of the vocal tract filter and the other is to change the efficiency of the energy transfer at the glottis. Our primary interest is whether the air sacs can optimize the sound radiation in either of these two ways.

The present paper is organized as follows. In Sec. II, the anatomical structures of air sacs that vary among species are reviewed to introduce the basic designs of air sacs. In Sec. III, the experimental setup, including the details of the physical models that simulate the vocal folds, vocal tracts, and air sacs, is described. A computational model is also outlined here. In Sec. IV, experimental results are reported. The final section is devoted to conclusions and discussions of the present study.

II. AIR SAC MORPHOLOGY, INFLATION, AND EXPERIMENTATION

Table I summarizes basic designs of air sacs and bulla and their positions relative to the vocal folds known from the existing studies (Bartels, 1905; Starck and Schneider, 1960; Hayama, 1970). In our experiments, we investigate two basic designs: an inflatable air sac and a non-inflatable bulla-like

cavity, both of which branch off from the supraglottal vocal tract. In the case of bulla, the cavity size is always kept constant during the phonation, which must occur with open mouth or nose. In contrast, the inflatable air sacs are not supported by a skeletal frame, but they have a flexible wall composed of high amounts of elastic and collagen fibers and, to some extent, muscle fibers. Air sac walls thus contain both passive and active components that facilitate collapse. The air sac, which is connected to the rest of the vocal tract cavity via a tube-like structure (either pairwise lateral ventricles or the single *ductus pneumaticus*), expands up to a certain size, ranging from a few milliliters up to several liters in large apes (Starck and Schneider, 1960; Nishimura *et al.*, 2007) and in muskox and reindeer (Frey *et al.*, 2006, 2007). For inflation, the oral and nasal cavities are closed and the animal impels air into the air sac (in a *Valsalva-like* maneuver).

The critical point of modeling the air sac seems to be whether phonation occurs with an open or a closed mouth. With an open mouth, the air sac might serve as some sort of a side branch. With a closed mouth, however, vocalization occurs with air flowing into the air sac, and sound is emitted through the soft air sac wall and skin, which serves as the main filter. In the siamang (*Symphalangus syndactylus*), two lateral ventricles merge ventrally from the larynx into a single large air sac at a ventral neck position. Anatomical measurements in an adult female siamang cadaver (provided by Dr. Nishimura at the Japan Primate Research Institute) revealed that the lateral ventricles have an inner diameter of 1–1.5 cm and a length of approximately 4 cm; these merge ventrally into the air sac, which can maximally expand to approximately 800–1000 ml. During the boom call production of siamangs, inflation of the air sac has been reported to be accompanied by closed lips or nares (Haimoff, 1981, Tembrock, 1974). When the lips are open, the air follows the least resistance and is expelled from the mouth and possibly from the nose. The air sac deflates accordingly. It should be noted that there exists some speculation that air sacs can inflate even with open lips. For instance, entrance to the air sac can be closed by some mechanism, which keeps the air sac inflated by separating it from the vocal tract (Schön-Ybarra, 1995). Another speculation is that exhalation pressure during open-mouth phonation alone is able to inflate the air sac (Frey *et al.*, 2007). These hypotheses are, however, not a subject of this study and have to await further investigations.

Little experimentation has been performed on the acoustic roles of the air sacs. Gautier (1971) conducted a fistulation surgery of the singular, ventrally located air sac in one male DeBrazza's monkey (*Cercopithecus neglectus*). Gautier observed an effect on the amplitude of higher harmonics but not on the location of any formant. Hilloowala and Lass (1978) also reported no effect on formant positions in rhesus monkeys (*Macaca mulatta*), which also possess a singular ventrally located air sac (Schneider, 1964). They surgically removed the air sac (without detailed description of the surgical procedure) in three males and found very similar frequency values before and after surgery. Although their report is somewhat confused with the terms "fundamental frequency" and "formant frequencies" (see Table I in Hilloowala and Lass, 1978), it is most likely that they report formant frequencies since their frequency values resemble the formant values reported for rhesus monkeys by Rendall *et al.* (1998).

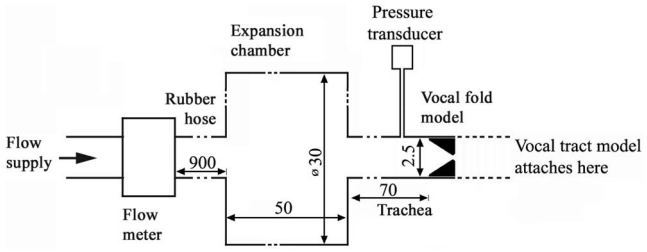

FIG. 1. Schematic of the subglottal parts of the experimental apparatus (numbers indicate distance in cm).

III. METHODS

Our experimental system is composed of a sophisticated, repeatedly tested vocal fold model, a simplified vocal tract model, and measurement devices (see Fig. 1). Details of each component are described sequentially in the following subsections.

A. Physical model of the sound source

A physical replica of the vocal fold system was constructed by using a silicone model of the vocal folds. The silicone model was a two-layer representation of the body-cover vocal fold layers. To simulate this body-cover composition, an earlier fabrication process for making one-layer (homogeneous) vocal fold models (Thomson *et al.*, 2005) was modified to include two layers. The procedure is illustrated in Fig. 2 and is as follows. Two three-dimensional computer-aided design (CAD) models were generated by using the commercial package PRO/ENGINEER. These two CAD models allowed the body and the cover to be made in series. For the first model geometry (model "A"), the dimensions of Scherer *et al.*, (2001) were used to define the vocal fold cover profile. The cross section was uniform in the dorsoventral direction. A base and two positioning guides were added to the vocal fold model to ensure proper orientation and cover thickness during fabrication. A second model (model "B") was generated by removing 2 mm of material from the vocal fold surface of model A. Rigid acrylonitrile butadiene styrene (ABS) physical models were made from the computer models using a Dimension™ rapid prototyping printer. Negative molds were made from these positive ABS models by using a two-part addition-cure silicon rubber (Smooth-Sil™ 950; this and other model-making materials referenced below are manufactured by Smooth-On Inc., Easton, PA, USA).

The two-layer vocal fold models were created by using a three-part addition-cure silicone compound (Ecoflex 0030), mixed with varying ratios of Ecoflex 0030 part A:Ecoflex 0030 part B:silicone thinner. The modulus of the cured silicone could be adjusted by varying the amount of thinner used. A ratio of 1:1:2 with a cured modulus of approximately 10 kPa was used for the body, and a ratio of 1:1:5 with a

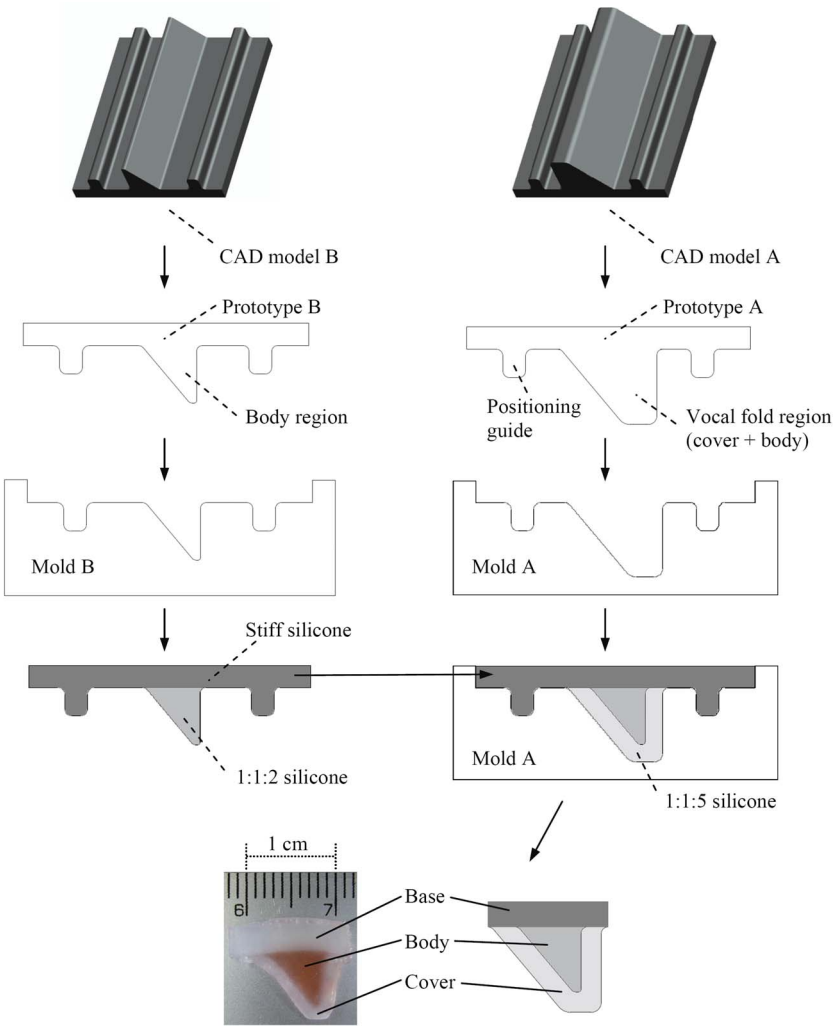

FIG. 2. (Color online) Schematic of the two-layer vocal fold fabrication process, including computer model generation, rapid prototyping, and casting of the different layers. Shown at lower left is an image of the cross section of the vocal fold model.

cured modulus of approximately 2 kPa was used for the cover. Modulus data were obtained by mounting cylindrical specimens (0.8 cm diameter and 6 cm long) of the given ratios in a tensile testing machine (Instron 3342) and strained at a rate of 1000 mm/min to 40% strain. Both ratios yielded materials with nearly linear stress-strain curves.

To fabricate the vocal folds, the body of the vocal fold model was created first. Mold B was sprayed with Universal Mold Release, and the 1:1:2 silicone mixture was poured into the mold so the liquid just filled the body area, leaving the base and guide volumes unfilled. After the body cured (~6 h), a very stiff silicone mixture (Dragon Skin) was poured into the mold to fill in the base and guides and allowed to cure for 2 h. The stiffness of the base was deemed to be sufficiently high so as to not vibrate with the vocal folds. A second mold release agent (Ease Release 800) was then sprayed into mold A. The 1:1:5 silicone mixture was poured into mold A. The silicone model cast using mold B was inserted into mold A. The cover was allowed to cure for 24 h and the entire model was removed from mold A. The portion of the base not directly under the vocal fold model was removed, and the model was then cut into two 1.7 cm lengths (for two symmetric vocal folds).

The dorsoventral and lateral surfaces of the model were attached to a 1.2 cm thick acrylic orifice plate using a liquid polyurethane adhesive (Pro Bond®, Elmer's Products, Inc.). Two such components were fastened together to form a single orifice plate, representing a full laryngeal configuration. The separation distance between the support plates was variable. The medial surface of the two folds was positioned to be in contact so that the glottis was closed when no airflow was applied. A uniform polyvinylchloride (PVC) tracheal tube (2.5 cm inner diameter) was connected upstream to an expansion chamber to simulate the subglottal system (Fig. 1).

A sweep tone experiment was conducted to stimulate the vocal tract models using an alternative sound source. To generate a well controlled sound, a horn driver (P-15F, UNIPEX) was utilized. The vocal tract models, which will be explained in detail in the following subsection, were attached to the horn driver. The sweep tone signals (pure tones whose frequency changes continuously from low to high frequency) were designed with the COOLEDIT 2000 software (Syntrillium Soft. Corp.).

B. Physical models of the vocal tract

The vocal tract models were made of PVC tubes. Four main configurations were designed as follows.

(1) As a reference configuration to study the effect of the air

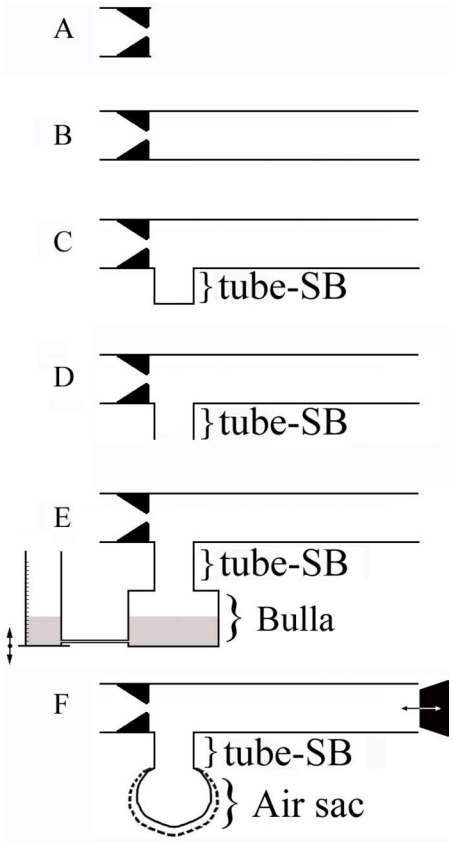

FIG. 3. (Color online) Schematic illustrations of the physical model configurations. Whereas no vocal tract is attached to the vocal fold physical model in (A) the following vocal tract models are attached to the vocal fold model in (B)–(F). (B) Single uniform tube as main vocal tract. (C) Main vocal tract with closed uniform tube side branch ("tube-SB"). (D) Main vocal tract with open tube-SB. (E) Main vocal tract with tube-SB and rigid cavity of variable volume ("bulla"). The volume is regulated by the water cavity level. (F) Main vocal tract with tube-SB and inflatable air sac ("air sac"). The inflation is regulated by "closing the mouth," i.e., inserting a plug to close the aperture of the vocal tract as if the lips are closed.

sac models, cases with no vocal tract (Fig. 3(A)) and with a main uniform tube vocal tract [25 mm inner diameter, 200 mm long; Fig. 3(B)] were considered. Note that in the situation without a vocal tract, there is no tube extending from the vocal folds.

(2) A tube-like side branch (hereafter "tube-SB") was attached to the main vocal tract [Figs. 3(C) and 3(D)]. The inner diameter was varied to examine its effect on the produced sound. Anatomically, this tube-SB may correspond to the connecting tube between the main vocal tract and a large actual air sac [lateral ventricle or *Ductus pneumaticus*; Table I(B)] or a very small air sac system alone. Both open and closed situations were considered. Acoustical measurements of the tube-SB are of particular importance for understanding the exclusive effect of the air sac models, which are attached to the vocal tract through the tube-SB.

(3) A non-inflatable bulla-like cavity was attached to the main vocal tract through the tube-SB [Fig. 3(E)]. Such a configuration exists, for example, in howler monkeys (subfamily *Alouattinae*). As a rigid non-inflatable cavity, a PVC cube ($100\times100\times100$ mm^2) was used. Variable water levels in the cavity, controlled through an outlet (3 mm) located at the lowest point of the cavity, enabled the investigation of different cavity volumes.

(4) An inflatable air sac was attached to the main vocal tract through the tube-SB [Fig. 3(F)]. Such a configuration exists, for example, in siamangs (*Symphalangus syndactylus*) and reindeers (*Rangifer tarandus*) (Frey *et al.*, 2007). An inflatable air sac was simulated by using a urinary bladder from a pig (*Sus scrofa f. domestica*) or a cow (*Bos taurus*). The bladder wall has similar wall characteristics as real air sacs. Namely, the wall contains high amounts of elastic and collagen fibers, which can be repeatedly stretched to its maximal size, i.e., few times larger than its relaxed state, without causing any damage (Ogura, 1915). Two situations were considered: closed mouth and open mouth, simulated by repeatedly opening and closing the open end (corresponding to the "mouth") of the main vocal tract with a plug. After inserting the plug, the sound persisted until the air sac model was inflated up to its maximum size.

Wall characteristics of the vocal tract model affect the formant bandwidth. A nonyielding wall (such as PVC) narrows the bandwidth and therefore might affect the overall sound pressure level. However, we are interested in the relative measurements and hence believe that the data can be compared in a fair manner. We expect the sound wave reflection on the water surface to be similar to that of the PVC surface.

C. Experimental setup

Vocal fold vibration is affected by resonances of the subglottal system (Titze and Story, 1997). A tracheal resonance (in particular, the negative reactance) in the vicinity of the fundamental frequency lowers the vocal fold efficiency (Titze, 2008), which is undesirable since a stable vocal source signal is needed here. Therefore, the length of the trachea tube was chosen based on the recommendations provided for a similar vocal fold system (Zhang *et al.*, 2006a). The vocal fold model was installed at the end of a 70 cm long PVC tube that models the trachea. An expansion chamber (inner cross-sectional diameter: 30 cm: length, 50 cm) was connected to the air flow supply through a 9 m long rubber hose so as to reduce possible flow noise from the air supply. This chamber served to simulate the change in cross-sectional area from the lungs to the primary bronchi (Ishizaka *et al.*, 1976). This design is quite similar to that tested in Zhang *et al.* (2006a), in which an ideal open-ended termination of the tracheal tube for frequencies above approximately 50 Hz was realized.

The sound generated from the physical model was recorded by using an omnidirectional microphone (Bruel and Kjær, type 4192; Nexus conditioning amplifier), located 20 cm from the sound source or from the vocal tract opening. The sound pressure level (SPL) was measured by a sound level meter (Bruel and Kjær, type 2250).

The subglottal pressure in the tracheal tube was monitored by using a pressure transducer (Differential pressure transducer, PDS 70GA, Kyowa, Japan) with a signal condi-

tioner (CDV 700A, Kyowa, Japan). The pressure transducer tube was mounted flush with the inner wall of the tracheal tube, 2 cm upstream of the vocal fold plates.

The volumetric flow rate through the glottis was measured by using a precision mass-flow meter (CMQ-V, Yamatake Corp., Japan) at the setup inlet. During the experiments, the flow rate was increased from zero to a certain maximum value in discrete increments. At each step, after the flow rate was changed, measurement was delayed for an interval of about 4–5 s in order to allow for flow field stabilization

All three signals were recorded as WAV files (16 bit resolution, 44 kHz sampling rate) on a digital recorder (R4 Edirol, Yamaha, Japan) or alternatively with a fast Fourier transform (FFT) analyzer (DL 750, Scopecorder, Yokogawa, Japan). Experiments were conducted in a sound attenuated room with controlled climate (22 °C, 75% humidity).

D. Computational simulation

In order to validate the experiments, acoustical characteristics of the physical model of the vocal tract were estimated by using the transmission line model (Flanagan, 1972; Sondhi and Schroeter, 1987). In this model, the vocal tract is represented by a series connection of many uniform tubes. Assuming plane-wave propagation, the input-output relationship of the pressure and the volume velocity at each tube is described in terms of a four terminal network. Acoustical characteristics of the vocal tract are obtained by computing the transmission characteristics of the cascaded connection of such lumped elements. This approach can be applied to estimate the transfer function of the vocal tract even in the presence of a side branch (Sondhi and Schroeter, 1987; Dang and Honda, 1996). As acoustic impedance parameters, standard values that assume a static wall condition were utilized (Flanagan, 1972). Radiation impedance at the open end of the vocal tract was determined according to Causse *et al.* (1984).

Configuration of the transmission line was constructed in accordance with the basic design of the physical model of the vocal tract including the air sac. A uniform tube was divided into a cascade connection of 1000 lumped elements. In order to match the first formant frequency, a corrected value of 19.8 cm was used as the length of the main vocal tract. All other parameters such as the length of the tube-SB, the tube diameter, and the cavity size were adopted from the original values of the physical model.

IV. RESULTS

A. Reference measurements

First, phonation threshold pressure (PTP), fundamental frequency (F0), and SPL values of the vocal fold physical model were measured without any vocal tract [Fig. 3(A)]. These measurements will serve as a reference to study the effect of different vocal tract configurations examined in the following experiments. PTP represents the minimum subglottal pressure required to initiate and sustain the vocal fold oscillations. Functionally, it provides an index for the vocal effort. PTP data are given in Table II.

TABLE II. Phonation threshold pressure (PTP) and flow rate at voice onset point for different vocal tract (VT) physical models. Values are averaged over three trial measures; the standard deviations in all cases are less than 0.5%. For uniform tube side branches (tube-SBs), two measurements are given: one for a closed side branch and the other for an open side branch.

VT model	PTP (cm H_2O	Flow rate (ml/s)
No VT	5.749	241
20 cm uniform tube	5.427	183
SB0.8ID closed/open tube-SB	5.734/5.646	216/216
SB1.3ID closed/open tube-SB	5.646/5.354	225/216
SB1.6ID closed/open tube-SB	5.719/5.383	233/216
SB1.3ID 1000 ml cavity	5.456	216
SB1.3ID 600 ml cavity	5.734	133
SB1.3ID 200 ml cavity	5.486	150

F0 of the vocal fold model ranged from 105 to 113 Hz for subglottal pressures between 6 and 20 cm H_2O. Stiffness and length of the vocal folds, which are considered as major causes to change F0, were kept constant. Figure 4 demonstrates the relationship between subglottal pressure and flow rate. The pressure-flow relationship is expected to be quadratic for an orifice with yielding walls, as is the case with the vocal folds (Alipour and Scherer, 1997).

The SPL ranged between 48 dBA (at PTP) and 69 dBA for subglottal pressures between 6 and 20 cm H_2O [Fig. 5(a)]. Next, as a basic design for all vocal tract models, only the main vocal tract model [Fig. 3(B)] was attached to the vocal fold model. Compared to the no vocal tract measurement, the main vocal tract model lowered PTP (Table II) and increased SPL for varied subglottal pressure [Fig. 5(a)]. Together with the no vocal tract measurement, this basic design is also utilized as a reference to study the relative effect of various air sac models added to the main vocal tract.

B. Side branch modeled as a uniform tube

In the case of a closed tube-SB connected to the main vocal tract [Fig. 3(C)], the SPL was further increased, where the tube-SB with an inner diameter of 0.4 or 0.8 cm provided the maximum SPL [Fig. 5(b)]. The closed tube-SB also low-

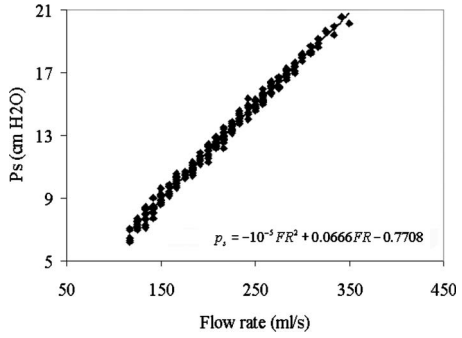

FIG. 4. Subglottal pressure–flow rate relationship of the vocal fold physical model during phonation. Data for the physical model with no vocal tract, data with a 20 cm uniform tube as the vocal tract, and data with all other tested vocal tract models are combined. Note that the pressure-flow rate relationship is identical for different vocal tracts.

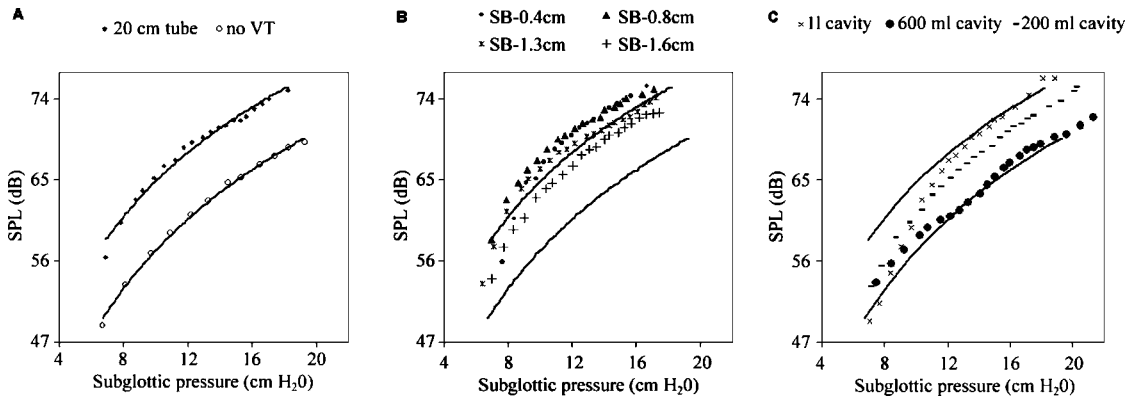

FIG. 5. Sound pressure level (SPL)—subglottal pressure relationship of the vocal fold physical model during phonation. (A) Without and with a 20 cm long tube as a vocal tract. (B) 20 cm long tube with tube-SB. Side branch is closed and mouth is open. Three side branches were tested; each has 5 cm length with different inner diameters (inner diameters, IDs: 0.4, 0.8, 1.3, and 1.6 cm). (C) 20 cm long tube with tube-SB. A rigid cavity of variable size is added to the SB tube. Results for three cavity sizes are shown (cavity sizes: 1.0, 0.6, and 0.2 l). The same trendlines as in Fig. 5(a) are drawn (lower line: physical model without vocal tract; upper line: 20 cm uniform tube vocal tract).

ered PTP below that of the basic design (Table II). The open tube-SB also showed PTP lower than the reference values (Table II).

Addition of these side branches introduces a pole/zero pair to the spectrum and consequently moves the neighboring formants (Table II). Among the four formants, the first formant is affected most strongly (12%–20%) (Table III). Depending on whether the tube-SB is closed or open, the additional pole/zero pair of a 5 cm tube-SB is located differently, at approximately 1.8 kHz for closed tube-SB and at approximately 3.6 kHz for open tube-SB (Fig. 6). Obviously, position of the pole/zero pair behaves as a quarter-wavelength resonator ($c/4L$, where c is the sound speed and L is the tube length) when the tube-SB is closed and as a half-wavelength resonator ($c/2L$) when the tube-SB is open.

To investigate whether the inner diameter of the tube-SB provides any limit on vocalization, we measured the relationship between the subglottal pressure and SPL for the open tube-SB models under the condition that the mouth was closed. The inner diameter of the side branch was varied (0.4, 0.8, 1.3, and 1.6 cm). Phonation with a side branch model with an inner diameter of 0.8 cm was possible only with a very high flow rate. No phonation was possible when the inner diameter was smaller than 0.8 cm. The results imply that there exists a minimum inner diameter required to maintain vocal fold vibrations, below which vocal fold vibration cannot be achieved (or at least demands an excessively high subglottal pressure). This may be because the increased flow resistance associated with the small tube-SB opening creates a high supraglottal pressure and thus makes the pressure difference between the subglottal and supraglottal systems insufficient for sustained vocal fold oscillation.

C. Non-inflatable bulla-like cavity

For the case of a rigid cavity attached to the main vocal tract, dependence of SPL on cavity size was studied [Fig.

TABLE III. Resonant frequencies obtained from measurements of different vocal tract (VT) models with closed and open side branches (SB: side branch; no SB: 20 cm long tube without side branch; cSB: closed side branch; oSB: open side branch; SB-0.8: side branch with inner diameter of 0.8 cm). The nth resonant frequency is denoted as Fn. The term "pole" stands for an additional resonant frequency introduced by a side branch or cavity. Values in parentheses show deviations in percentage from the resonant frequencies of the 20 cm long tube without a side branch.

VT model	pole (Hz)	F1 (Hz)	F2 (Hz)	pole (Hz)	F3 (Hz)	F4 (Hz)
No SB		380	1170		1990	2560
cSB-0.4		320 (−15.7)	1250 (+6.8)	1640	1890 (−5)	2460 (−3.9)
cSB-0.6		300 (−21)	1250 (+6.8)	1630	1890 (−5)	2450 (−4.3)
cSB-0.8		290 (−23.7)	1140 (−2.5)	1620	1820 (−8.5)	2400 (−6.3)
cSB-1.3		300 (−21)	1160 (−1)	1630	1880 (−5.5)	2500 (−2.3)
cSB-1.6		290 (−23.7)	1115 (−4.7)	1650	1900 (−4.5)	2500 (−2.3)
oSB-0.4		310 (−18.7)	1230 (+5.1)		1875 (−5.8)	2465 (−3.7)
oSB-0.6		320 (−15.8)	1260 (+7.7)		1890 (−5)	2465 (−3.7)
oSB-0.8		370 (−2.6)	1220 (+4.3)		1820 (−3.7)	2410 (−5.8)
oSB-1.3		380 (0)	1270 (+8.5)		1900 (−4.5)	2410 (−5.8)
oSB-1.6		380 (0)	1260 (+7.7)		1890 (−5)	2410 (−5.8)
cav 100 ml	115	660 (+73.7)	1320 (+12.8)		2110 (+11.6)	3030 (+18.3)
cav 500 ml	78	650 (+71)	1310 (+12)		2110 (+11.6)	3030 (+18.3)

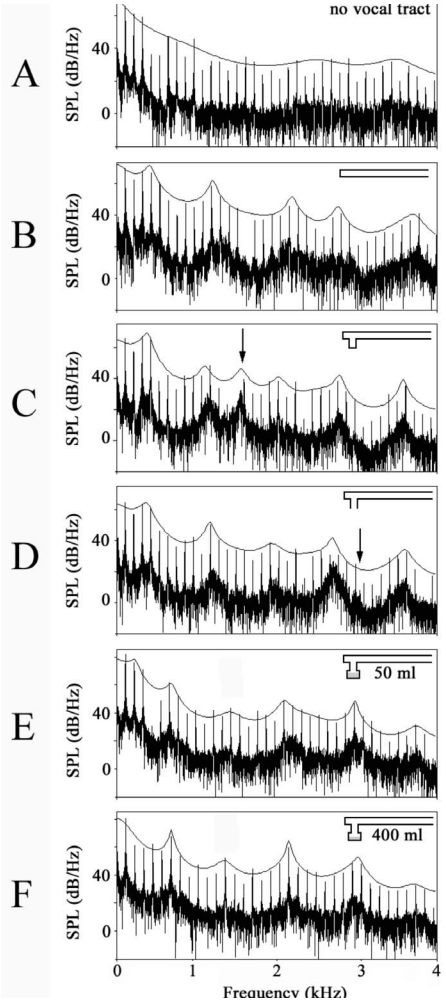

FIG. 6. Spectra and linear predictive coding (LPC) envelopes for different physical models. The vocal tract model is schematized in the top right corner of each spectrum. (A) No vocal tract [as in Fig. 3(A)], (B) uniform tube vocal tract [as in Fig. 3(B)], (C) uniform tube vocal tract with closed side branch [as in Fig. 3(C)] (arrow indicates additional pole), (D) uniform tube vocal tract with closed side branch [as in Fig. 3(D)], (E) uniform tube vocal tract with rigid cavity (50 ml volume) [as in Fig. 3(E)], (F) uniform tube vocal tract with large rigid cavity (400 ml volume) [as in Fig. 3(E)]. Note that the additional zero in the vocal tract with the open side branch overlaps with the fourth formant [arrow in panel (D)].

5(c)]. Compared to the 600 ml cavity, the 200 ml cavity gives higher SPL values, which can even exceed those of the basic design for high subglottic pressures.

It should be noted that Figs. 5(a)–5(c) reveal an interesting effect of the tube-SB as well as of the rigid cavity. Namely, the slope of the SPL-subglottal pressure relationship becomes much steeper compared to the reference measurement or the basic design. For instance, increasing the subglottic pressure from 8 to 16 cm H_2O caused an increase in SPL by 12 dB (from 54 to 66 dB) for the reference measurement [Fig. 5(a)]; a comparable amount of increase in SPL was also observed for the basic design (from 60 to 73 dB). However, in the case of the attachment of a tube-SB [Fig. 5(b), 1.3 cm diameter] or a 1 l rigid cavity [Fig. 5(c)], the same increase in subglottic pressure (from 8 to 16 cm H_2O) resulted in an increase in SPL of 20 dB (from 53 to 73 dB). This implies that within the same range of subglottic pressure, dynamic range of the sound emission can be widened by introducing the side branch or the rigid cavity.

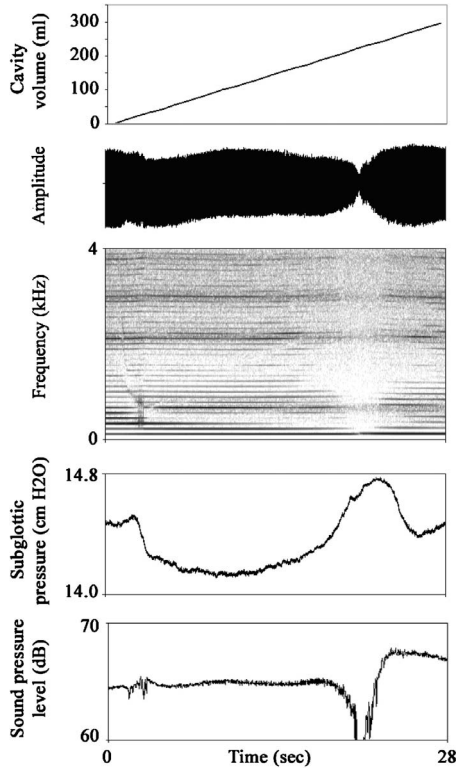

FIG. 7. Experiment with a continuous change in rigid cavity volume attached to the main vocal tract. During the experiment, the cavity was gradually filled with water. Note the aphonic regime when the resonant frequency crossed F0. Phonation started again slowly after the resonance moved beyond from F0. The SPL reached a higher level after the crossing.

To study the dependence of SPL on the rigid cavity volume in more detail, another experiment was conducted in which the cavity size was gradually changed from 0 to 300 ml (Fig. 7). At a cavity size of approximately 210 ml, the physical model ceased to vibrate. The corresponding PTP was too high to measure at this point (Fig. 8). This phenomenon is due the following nonlinear source-filter interaction. The pole/zero pair introduced by the rigid cavity depends on the cavity size (Table III). A small cavity introduces a pole/zero pair around 150 Hz. As the cavity size is increased (from 0 to 350 ml in Fig. 7), the pole/zero pair decreases and eventually crosses the fundamental frequency of the vocal fold model (F0=110 Hz) at a cavity volume of

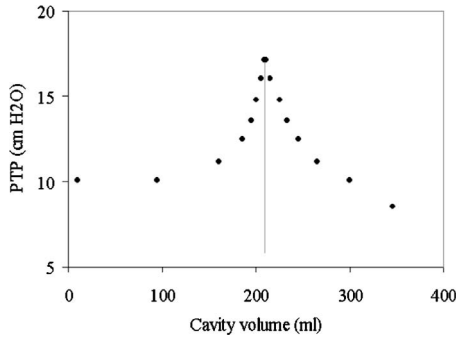

FIG. 8. Dependence of phonation threshold pressure (PTP) on the volume of the rigid cavity attached to the main vocal tract.

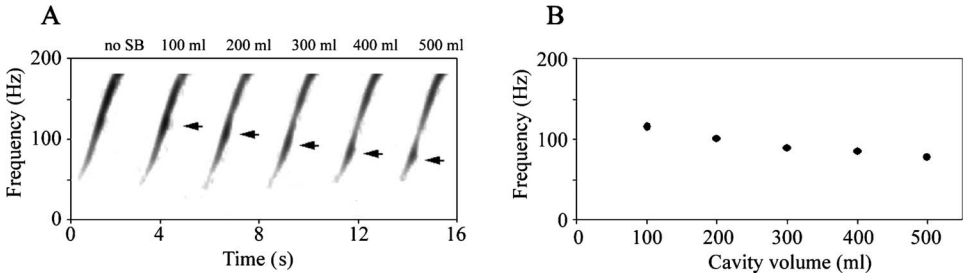

FIG. 9. (A) Spectrograms of the vocal tract model stimulated by sweep tones ranging from 90 to 150 Hz. As indicated above each spectrogram, the cavity volume was changed from 0 to 500 ml. The arrows indicate the location of a pole (vocal tract resonance peak) introduced by the cavity; "no SB" stands for no side branch. (B) Pole-cavity size relationship. For each sweeptone, an averaged FFT was calculated. The pole shows up as a peak in the FFT spectrum.

approximately 210 ml. The interaction between the sound source and the vocal tract becomes very strong and inhibits the model vibrations. Once the pole/zero pair passes away from the fundamental frequency, the source-tract interaction weakens and the vocal folds resume oscillation. The interaction effect is different on both sides of this phonation stop. The SPL was higher after the phonation stop, implying that phonation is facilitated when F0 is larger than the pole/zero pair. Before and after the phonation stop, the subglottal pressure changed by approximately 1 cm H_2O. F0 also changed from 110 to 120 Hz.

To locate a precise value of the pole/zero pair and to study its dependence on the cavity size, the transfer function of the vocal tract model was determined by a sweep tone experiment (Fig. 9). In this experiment, the vocal tract model was stimulated by a horn driver with a variable frequency ranging from 90 to 150 Hz. The cavity size was changed from 0 to 500 ml with an increment of 100 ml. The spectral analysis of the output sound shows that the pole/zero location is lowered as the cavity size is increased (Fig. 9). This is in good agreement with the computational simulation (see Sec. IV E below).

D. Inflatable air sac

To study an inflatable air sac added to the side branch [Fig. 3(F)], two distinct situations were considered: closed mouth (plug inserted) and open mouth (plug not inserted). For a given air flow rate (for instance, 34 l/min), the SPL measured with a closed mouth was approximately 59 dB for the pig urinary bladder (approximately 500 ml maximum inflation, approximately 10 cm diameter) and 68 dB for the cow urinary bladder (approximately 5 l maximum inflation, approximately 20 cm diameter).

After the plug was released from the mouth, the air sac was deflated. Gravity of the air sac, recoiling properties of the elastic tissue, and the air exiting through the open mouth caused a complete collapse. Partial inflation of the air sac, which is sometimes observed *in vivo*, was not simulated in the present experiment.

Spectral analysis of the sound emitted from the inflating air sac model indicates one major resonance in the area of 250 Hz [Fig. 10(a), bottom]. Higher harmonics are diminished. This is in good agreement with the siamang boom call, which has a major peak at the fundamental frequency between 200 and 300 Hz and only weak energy at higher harmonics [Fig. 10(b), bottom].

As shown in Figure 10(b), parts of the siamang song are characterized by a stereotypic repetition of the boom calls (arrow 2) and the scream calls (arrow 1). The fundamental frequency jumps from approximately 250 Hz in the boom call to approximately 800 Hz in the scream call. In the physical model, a relatively small shift in the fundamental frequency (up to 15 Hz) was observed when the closed mouth was quickly opened. This sudden mouth opening induces a situation with high glottal flow and high transglottal pressure, which is known to cause F0 changes in human voice (Hsiao *et al.*, 1994). Since our physical model has rather a small F0 variability (presumably primarily due to the linear stress-strain material response), the observed F0 jump was not as large as the real siamang vocalization.

The deflation process was very interesting. Using a large cow bladder, the deflation lasted for about 5–10 s. During this deflation process, the spectral characteristics of the voice initially resembled those of an open tube-SB and then later those of a closed tube-SB. This can be confirmed in the sweep tone data shown in Fig. 11. During deflation, the pole/zero pairs (arrow 2) are located at the same positions as the open tube-SB ($nc/2L$, where n is an integer). In the collapsed configuration, the pole/zero pairs (arrows 3 and 4) are located at the same positions as the closed tube-SB [$(2n-1)c/4L$]. In addition, we found in some cases a zero frequency in the range between 100 and 200 Hz. This is very similar to the pole/zero location of the non-inflatable cavity (see Sec. IV C). In contrast to the non-inflatable cavity, we never observed a phonation stop during the experiment.

We note that the vocal fold physical model was worn out rather quickly during this experiment. It is speculated that the repeated crossing of the zero frequency through F0 caused a higher stress in the vocal fold model and hence accelerated structural deterioration.

E. Computational simulation

To examine the acoustical characteristics of the physical model of the vocal tract, computational simulations were carried out to estimate the transfer functions for various configurations of the vocal tract. Figure 12 compares the transfer function of the main vocal tract (basic design) with that of the vocal tract with open or closed side branch. The open

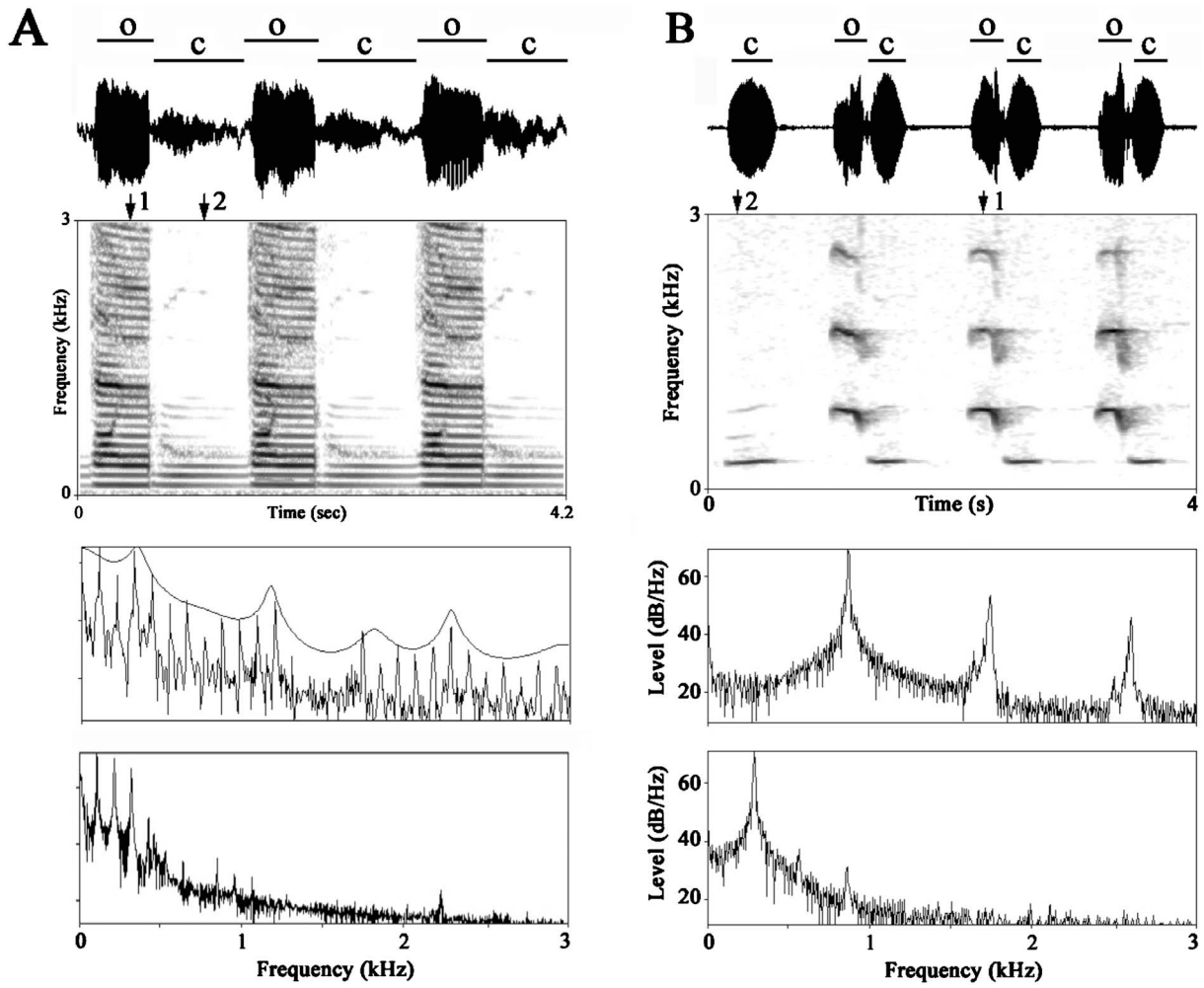

FIG. 10. Sound wave form (top panels), its spectrogram (second panels), and two spectra (third and fourth panels) from vocal fold physical model with inflatable bladder cavity (A) and from a siamang vocalization (B). Spectra are calculated from 50 ms segments indicated with arrows "1" and "2." Note that the formant structure of the open-mouth vocalization with a collapsed air sac [vocal tract model as schematized in Fig. 3(F)] resembles that of the closed side branch [see Fig. 6(c)]. "o"—open mouth or tube; "c"—closed mouth or tube.

side branch produced a pole/zero pair at approximately 3.6 kHz, whereas the closed side branch produced a pole-zero pair at approximately 1.8 kHz. The neighboring formants were moved accordingly. These results are in good agreement with the experiments summarized in Table III.

The transfer functions were also computed for the vocal tract with a rigid cavity [Fig. 13(a)]. The rigid cavity produced a pole/zero pair around 150 Hz. Location of this pair is sensitive to the cavity volume, where the dependence of the pole/zero frequencies on the cavity size is drawn in Fig. 13(b). Again, this agrees quite well with the experimental curve shown in Fig. 9. The computational model also identifies an additional pole/zero pair in the midfrequency range [Fig. 13(a)]. This pair at around 3.6 kHz might be related to the one computed for the open side branch in Fig. 12.

The experimentally obtained PTP curve shown in Fig. 8 can also be predicted by the computational model. Titze (1988) determined theoretically that PTP is dependent not only on the vocal fold parameters but also on the vocal tract input impedance. Since the input impedance corresponds to the acoustic load imposed on the vocal source, higher input impedance of the vocal tract requires a higher subglottic pressure to induce vocal fold oscillations. In Fig. 14, input impedance was drawn for the vocal tract with variable cavity size. Assuming that the input comes from the vocal source, the input impedance was computed by the convolution of the transfer function of the vocal tract with the power spectrum of the sound source signal measured from the vocal fold model (without any vocal tract). Although the computed input impedance has a peak at approximately 300 ml, which is somewhat different from the experiment, the qualitative structure of the experimental curve of Fig. 8, including the asymmetry between the left and right sides of the peak, is well reproduced by the computational model. One of the reasons for the inconsistency may be the bidirectional source-tract interaction, which is not considered in the present computation but which exists in the real experiment. The obtained peak frequency can be utilized to predict the phonation stop induced by the strong source-tract interaction of Fig. 7.

V. DISCUSSIONS

Experiments and computer simulations were carried out to investigate the questions on whether mammalian air sacs

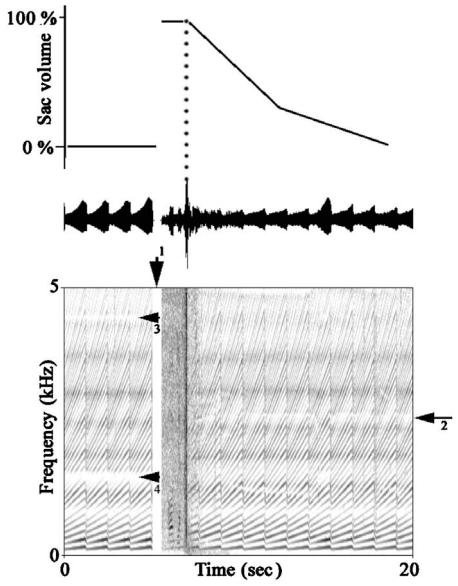

FIG. 11. Spectrogram of sweep tone (90–150 Hz) transmitted through a 20 cm tube vocal tract system with an inflatable air sac. Two situations are examined; first, a completely collapsed sac, up to about 5 s (arrow 1); second, continual deflation from air sac volume of about 5 l to complete deflation (from arrow 1 to end of sound). Arrow 2 indicates the location of zero frequency around 2500 Hz (which corresponds to zero frequency produced by the open side branch of 5 cm length). Arrows 3 and 4 indicate locations of zero frequency at about 1500 and 4500 Hz (which correspond to zero frequency produced by the closed side branch of 5 cm length). Note that during the deflation process (approximately between 9 and 14 s), another zero-like structure appears at approximately 1400 Hz, decreasing to 1300 Hz.

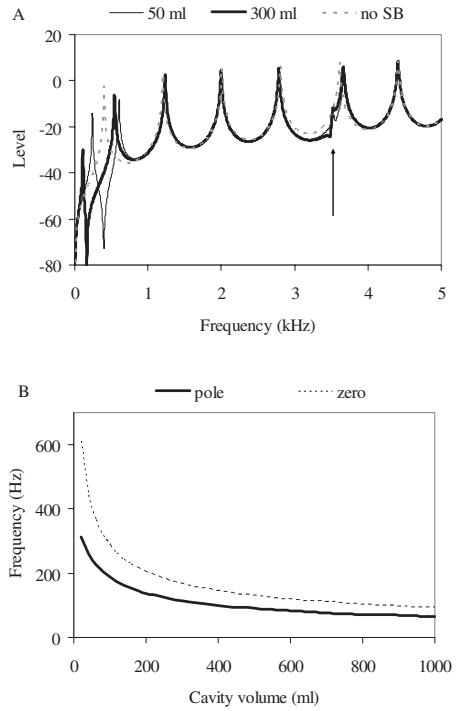

FIG. 13. (A) Transfer functions obtained by the computational model that simulates two rigid cavities (50 and 300 ml volume). Note that the first formant is affected by the pole/zero pair located in the low frequency range. The uniform tube that connects the main vocal tract and the rigid cavity introduces another pole/zero pair around 3.6 kHz (arrow), which resembles that of an open side branch. (B) Dependence of the pole/zero location on the cavity volume of the rigid cavity attached to the main vocal tract.

or bulla-like cavities affect formant position and/or SPL. Four vocal tract configurations were investigated: (1) a uniform tube-like main vocal tract without any attachment, (2) a vocal tract with open or closed uniform tube-SB, (3) a vocal tract with a tube-SB plus a rigid cavity air sac, and (4) a vocal tract with a tube-SB plus an inflatable air sac.

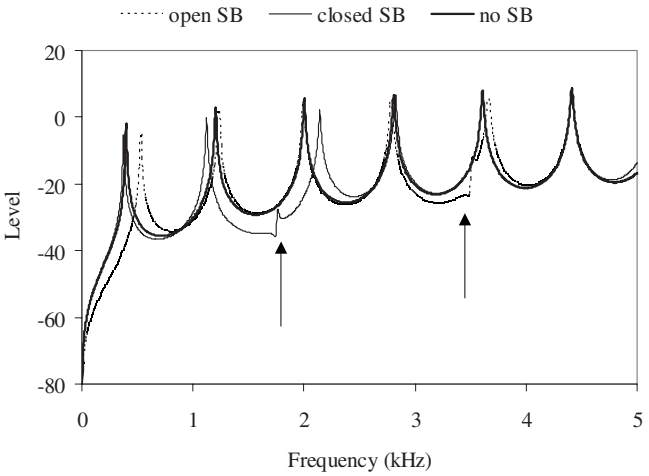

FIG. 12. Transfer functions obtained by the computational model that simulates the vocal tract without a side branch and the vocal tract with an open or a closed uniform side branch. Depending on whether the side branch is closed or open, an additional pole/zero pair of a 5 cm uniform side branch is located differently, at approximately 1.8 kHz for closed side branch (left arrow) and at approximately 3.6 kHz for open side branch (right arrow). The position of the pole/zero pair behaves according to a quarter-wavelength resonator when the side branch is closed and to a half-wavelength resonator when the side branch is open.

A uniform tube shapes the sound source signal in an expected way, i.e., formants at uneven multiples of $c/4L$ are present [Fig. 6(b)]. The 5 cm long tube-SB introduced a pole/zero pair in the midfrequency range (1.5–4.0 kHz, the exact position depends on whether the side branch is open or closed or on the length or the inner diameter of the side branch) (Dang and Honda, 1997; Dang *et al.*, 1994, 1998). As a consequence, all neighboring formants were shifted somewhat, where the first formant was most strongly shifted up to 25% [Figs. 6(c) and 6(d)]. The second and higher resonances were changed by not more than 10%.

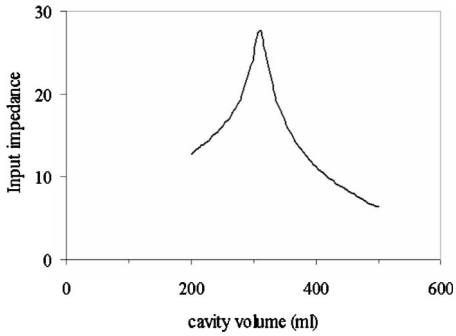

FIG. 14. Dependence of the input impedance of the vocal tract with a rigid cavity on the cavity volume. As an input signal to the vocal tract, reference sound measurement from the vocal fold physical model without a vocal tract is utilized. Since the input impedance is one of the key parameters that determine the PTP, the curve can be compared to the curve in Fig. 9.

A rigid cavity added to the tube-SB, which models the bulla hyoidea most prominent in *Allouatinae*, added a pole/zero pair in the range between 80 and 150 Hz. Again, the neighboring formants were shifted somewhat, where the first formant was shifted by up to 70% [Figs. 6(e) and 6(f)]. Finally, the inflatable air sac model added to a tube-SB was studied. In the case of a closed mouth, the air sac represents the main vocal tract, which gives rise to one major resonance below 400 Hz. On the other hand, if the mouth is open, the spectral characteristics depend on the degree of air sac inflation, which resembles either a rigid cavity or a closed side branch.

All air sac models mentioned above also have an effect on the SPL. Because of the nonlinear source-tract interaction, dependence of the SPL on the air sac configuration is hard to be systematically understood. Nevertheless, our experimental study showed that compared to the basic vocal tract configuration having no side branch, certain air sac configurations can enhance the SPL, for example, the closed tube-SB with inner diameter less than 1 cm and the rigid cavity of relatively small volume.

The present experiments also showed that the tube-SB and the rigid cavity models can widen the dynamic range of the sound emission within the same subglottal pressure range [Figs. 5 and 7]. Such physiological configurations would extend the vocal efficiency not only at higher pressures that make the sound louder but also at lower pressures. The lower levels are useful for "private" talk, in which the sender wishes to exclude unwanted listeners (eavesdroppers). To keep the voice "intelligible" in such private talk, the sound source must maintain its stable oscillations even at lower pressure levels.

A. Inflatable air sac cavity

Our inflatable air sac model suggests the presence of a single resonant frequency when the mouth is closed. Since there is no sound radiation from the mouth, the phonation occurs presumably through the air sac and skin walls. According to our experiment, when small air sac models (pig urinary bladder) were utilized, the closed mouth configuration gave a lower SPL than that of the open mouth. When large air sac models (cow urinary bladder) were utilized, the SPL values were comparable between the open and closed mouth configurations. The latter case is similar to siamang calls, which exhibit similar amplitude between scream call and boom call.

In siamang boom call, F0 is located around 250 Hz, which is close to the air sac resonance. Higher harmonics are mostly filtered out but can be observed sometimes [Fig. 10(b), bottom panel]. This F0 matching with the resonant frequency implies more efficient transmission of the sound energy than a mistuned system like our model, in which F0 (110 Hz) is more than an octave below the resonance (approximately 250 Hz) of the air sac [Fig. 10(a), bottom panel].

Phonation with closed oral and nasal cavities and into inflatable cavities has also been discussed in other contexts. For instance, ring doves (*Streptopelia risoria*) vocalize with a closed beak. They expel the air and phonate into an inflatable esophagus during their coo-call production (Riede *et al.*, 2004). The inflated esophagus of the dove is likely to act as an amplifier as well as a bandpass filter (Fletcher *et al.*, 2004). A modeling study based on the Helmholtz resonator clarified that the resonant frequency of the inflated cavity is determined mainly by the wall characteristics of the cavity and it is rather insensitive to the cavity size (Fletcher *et al.*, 2004). This enables the F0-resonance matching for a variable cavity size. Applying this idea to our inflatable cavity, we speculate that the air sac resonance is mainly determined by the wall characteristics of the bladder used in the experiment.

A large group of nonhuman primates with small air sac cavities (100 ml and less) remains to be discussed. Phonation into such a small cavity with a closed mouth situation probably plays no role, since the capacity of the air impelled into the sac is limited. Moreover, the connecting tube to those air sacs is relatively narrow ("less" than 10 mm in diameter) (Starck and Schneider, 1960). This narrow tube connection is likely to be insufficient to produce vocalizations with a closed mouth, as our side branch experiments with small inner diameter suggest. On the other hand, when the mouth is open, our closed tube-SB experiments demonstrate that there can be a positive effect of the narrow tube-SB on the SPL. These structures can thus help raise the SPL above that of the basic vocal tract configuration. This kind of sound amplification can be expected if we may regard the small air sacs as a closed tube-SB. A quarter-wavelength resonance should be detected in the spectrum.

B. Formant deviation

It seems unlikely that an air sac (either inflatable or noninflatable) alone is able to explain the severe formant deviations observed in nonhuman vocalizations. Our noninflatable air sac models changed the first formant by up to 72%, the second formant by up to 20%, and the higher formants less. In Diana monkey alarm calls, however, the first and second formants are deviated from those of the uniform vocal tract by 23% and 60%, respectively (Riede *et al.*, 2005). The severe deviation observed in the second formant seems unlikely to be produced by the air sac. Although the air sacs have been suggested to be the main reason for this effect (Lieberman, 2006), discontinuities in the nonhuman vocal tract are more likely to be the candidate (Riede *et al.*, 2006).

C. Nonlinear source-filter interaction and the efficiency of the sound source

The physical model of the vocal folds experienced lower PTP and increased SPL after the main vocal tract was added. The vocal tract models with side branches or rigid cavities also lowered PTP. Certain side branch models can decrease PTP even below that of the basic vocal tract configuration. This phenomenon is due to the nonlinear source-filter interaction as explained by theoretical studies, which state that PTP decreases as the inertive impedance of the vocal tract increases (Titze, 2004a, 2004b) and which have been confirmed in other physical model experiments (Chan and Titze,

2006). Given a similar basic design of the vocal fold source and the vocal tract filter, it is reasonable to assume the nonlinear source-filter characteristics also in nonhuman mammals, just as in humans performing specific vocal exercises (Titze *et al.*, 2008).

Of particular interest is the region where F0 and vocal tract resonance (pole/zero pair) are in close proximity. Near this region, the vocal tract impedance is quite variable and may produce dramatic changes in the source-filter interaction. This has been most impressively demonstrated here by the low frequency pole/zero pair introduced by the cavity. As the cavity resonance crosses F0, the interaction becomes so strong that the vocal folds stop vibrating. The experimental results suggest that the effect of the source-filter interaction is asymmetric about both sides of the filter resonance. For instance, SPL was higher when F0 was on the right side of the resonance than when it was on the left side (Fig. 7). This suggests that a perfect matching between vocal tract resonance and F0 is not as desirable, as previously considered (e.g., Riede *et al.*, 2006), but F0 slightly off the resonance is more beneficial for energy feedback to the sound source (Titze, 2008).

Another positive effect on SPL was found in the experiments, possibly due to the additional pole/zero pair that shifted the other formants in such a way that the fundamental frequency was located in a more beneficial range of the vocal tract impedance. More efficient energy feedback from the vocal tract to the sound source can make the sound overall louder. This provides another possibility of increasing the sound level, besides the optimal radiation theory.

In general, nonlinear source-filter interaction can be induced by various resonant frequencies, which are not necessarily due to the air sacs. However, compared to a simple situation of a vocal tract having no air sacs, the presence of the air sac gives rise to additional pole/zero pairs that may produce more areas of enhanced source production as well as more areas of source instabilities.

Howler monkeys (*Allouatinae*) are the most prominent examples with a large *Bulla hyoidea* (Starck and Schneider, 1960; Schön-Ybarra, 1995). It is interesting to note that howler monkey vocalization is characterized by a source spectrum with many bifurcations (Whitehead, 1995). Our preliminary investigation of those vocal utterances (unpublished data) suggests that source instabilities occur more often in the howler monkeys than, for instance, in Japanese macaques or chimpanzees which have an inflatable cavity (Riede *et al.*, 1997, 2004b, 2007) or in dogs and cats which possess no air sacs (Riede *et al.*, 1997, 2000).

D. Advantages and disadvantages of possessing an air sac for vocal communication

To summarize our discussions, we have recognized two acoustic effects of the air sacs. First, the air sac can increase the dynamic range of the sound emission within a given subglottic pressure range, both at the upper and lower limits. Second, vocal variability can be increased in the following three ways. (A) In the case of inflatable air sacs, open and closed mouth situations enable a switch between different radiation and filtering mechanisms. (B) In the case of noninflatable air sacs, they extend the frequency range, in which voice instability and nonlinear phenomena are induced by source-tract coupling. This provides additional vocal variability. (C) The formants of the main vocal tract are shifted by a pole/zero pair introduced by air sac. Our experiments suggest that such formant shifts are most prominent on the first formant and less on the higher formants.

The latter point leads to the question of why humans do not possess any major extensions of the vocal tract. All four great apes do possess air sacs, but not humans. Obviously, humans developed a much more convenient way to alter formants, i.e., speech articulation through varying cross-sectional vocal tract area. For predominantly close-range communications, such as human speech, of particular importance are neither loudness nor vocal variability but a stable sound source and well articulated formants. In human speech, the fundamental frequency ranges between 100 and 300 Hz (average male F0, 110 Hz; average female F0, 220 Hz). Among normal phonations of various vowels, the ones which possess the first formant closest to F0 are English vowels /i/ (first formant, 300–400 Hz) and /u/ (first formant, 280–350 Hz). The first formants in both cases are still far away from F0 in both sexes. If a pole/zero pair close to the F0 range was created by an air sac, this would interfere with F0 and could cause more voice instabilities in human voice. In this sense, the absence of air sacs in humans may be essential for maintaining a stable sound source.

ACKNOWLEDGMENTS

T.R. was supported by a short-term Postdoctoral Fellowship for North American and European Researchers by the Japan Society for the Promotion of Science (JSPS) (PE06030) and by a Postdoctoral fellowship by the "Deutsche Akademie der Naturforscher Leopoldina" (BMBF-LPD 9901/8-127). I.T.T. was supported by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan. J.B.M. and S.L.T. were supported in part by NIH Grant No. NIDCD R01 DC05788. For technical support, we thank Dr. Ishikawa. For many helpful discussions, we are grateful to Professor Dang and Dr. Fujita.

Alipour, F., and Scherer, R. C. (**1997**). "Effects of oscillation of a mechanical hemilarynx model on mean transglottal pressures and flows," J. Acoust. Soc. Am. **110**, 1562–1569.

Bartels, P. (**1905**). "Über die Nebenräume der Kehlkopfhöhle," Z. Morphol. Anthropol. **8**, 11–61.

Causse, R., Kergomard, J., and Lurton, X. (**1984**). "Input impedance of brass musical instruments," J. Acoust. Soc. Am. **75**, 241–254.

Chan, R. W., Titze, I. R., and Titze, M. R. (**1997**). "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **101**, 3722–3727.

Chan, R. W., and Titze, I. R. (**2006**). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362.

Dang, J., Honda, K., and Suzuki, H. (**1994**). "Morphological and acoustical analysis of the nasal and paranasal cavities," J. Acoust. Soc. Am. **96**, 2088–2100.

Dang, J., and Honda, K. (**1996**). "Acoustic characteristic of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation," J. Acoust. Soc. Am. **100**, 3374–3383.

Dang, J., and Honda, K. (**1997**). "Acoustic characteristics of the piriform fossain models and humans," J. Acoust. Soc. Am. **101**, 456–465.

Dang, J., Shadle, C. H., Kawanishi, Y., Honda, K., and Suzuki, H. (**1998**).

"An experimental study of the open end correction coefficient for side branches within an acoustic tube," J. Acoust. Soc. Am. **104**, 1075–1084.
Fant, G. (**1960**). "Acoustic Theory of Speech Production," 2nd ed. (Mouton, The Hague, The Netherlands).
Flanagan, J. L. (**1972**). *Speech Analysis, Synthesis and Perception* (Springer, Berlin).
Fletcher, N. H., and Tarnopolsky, A. (**1999**). "Acoustics of the avian vocal tract," J. Acoust. Soc. Am. **105**, 35–49.
Fletcher, N., Riede, T., Beckers, G. J. L., and Suthers, R. A. (**2004**). "Vocal tract filtering and the 'coo' of doves," J. Acoust. Soc. Am. **116**, 3750–3756.
Frey, R., Gebler, A., and Fritsch, G. (**2006**). "Arctic roars—laryngeal anatomy and vocalization of the muskox (*Ovibus moschatus* Zimmermann, 1780, Bovidae)," J. Zool. **268**, 433–448.
Frey, R., Gebler, A., Fritsch, G., Nygren, K., and Weissengruber, G. E. (**2007**). "Nordic rattle—the hoarse phonation and the inflatable laryngeal air sac of reindeer (*Rangifer tarandus*)," J. Anat. **210**, 131–159.
Gautier, J. P. (**1971**). "Etude morphologique et fonctionnelle des annexes extra-laryngées des cercopithecinae; liason avec les cris d'espacement," Biol. Gabon. **7**, 229–267.
Haimoff, E. H. (**1981**). "Video analysis of Siamang (*Hylobates syndactylus*) songs," Behaviour **76**, 128–151.
Hayama, S. (**1970**). "The Saccus larynges in Primates," Journal of the Anthropological Society of Nippon. **78**, 274–298.
Hill, W. C. O. and Booth, A. H. (**1957**). "Voice and Larynx in African and Asian Colobidae," J. Bombay Natural His. Soc. **54**, 309–321.
Hillowala, R. A., and Lass, N. J. (**1978**). "Spectrographic analysis of air sac resonance in rhesus monkeys," Am. J. Phys. Anthropol. **49**, 129–132.
Hirano, M., and Kakita, Y. (**1985**). "Cover-body theory of vocal fold vibration," in *Speech Science: Physiological Aspects*, edited by R. Daniloff (College-Hill, San Diego, CA), pp 1–46.
Hsiao, T., Solomon, N. P., Luschei, E. S., Titze, I. R., Kang, L., Fu, T., and Hsu, M. (**1994**). "Effect of subglottic pressure on fundamental frequency of the canine larynx with active muscle tensions," Ann. Otol. Rhinol. Laryngol. **103**, 817–821.
Kurita, S., Nagata, K., and Hirano, M. (**1983**). "A. comparative study of the layer structure of the vocal fold," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 3–21.
Ishizaka, K., Matsudaira, M., and Kaneko, T. (**1976**). "Input acoustic-impedance measurement of the subglottal system," J. Acoust. Soc. Am. **60**, 190–197.
Lieberman, P. (**2006**). "Limits on tongue deformation—Diana monkey formants and the impossible vocal tract shapes proposed by Riede *et al.* (2005)," J. Hum. Evol. **50**, 219–221.
Neubauer, J., Zhang, Z., Miraghaie, R., and Berry, D. A. (**2007**). "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118.
Nishimura, T., Mikami, A., Suzuki, J., and Matsuzawa, T. (**2007**). "Development of the laryngeal air sac in chimpanzees," International Journal of Primatology **28**, 483–492.
Ogura, Y. (**1915**). "Beitraege zur Kenntnis des Kehlsackes des Rentieres," J. Coll. Agr. Tohoku Imp. University, Sapporo **6**, 151–155.
Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). "The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations," J. Acoust. Soc. Am. **103**, 602–614.
Riede, T., Herzel, H., Mehwald, D., Seidner, W., Trumler, E., Böhme, G., and Tembrock, G. (**2000**). "Nonlinear phenomena and their anatomical basis in the natural howling of a female dog-wolf breed," J. Acoust. Soc. Am. **108**, 1435–1442.
Riede, T., Arcadi, A. C., and Owren, M. (**2004b**). "Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos," Am. J. Primatol **64**, 277–291.
Riede, T. and Titze, I. R. (**2008**). "Vocal fold elasticity of the Rocky Mountain elk (*Cervus elaphus nelsoni*) — producing high fundamental frequency vocalization with a very long vocal fold," J. Exp. Biol. (in press).
Riede, T., Arcadi, A. C., and Owren, M. J. (**2007**). "Nonlinear acoustics in pant hoots and screams of common chimpanzees (*Pan troglodytes*). Vocalizing at the edge," J. Acoust. Soc. Am. **121**, 1758–1767.
Riede, T., Beckers, G. J. L., Blevins, W., and Suthers, R. A. (**2004**). "Inflation of the esophagus and vocal tract filtering in ring doves," J. Exp. Biol. **207**, 4025–4036.
Riede, T., Bronson, E., Hatzikirou, H., and Zuberbühler, K. (**2005**). "Vocal production mechanisms in a non-human primate: Morphological data and a model," J. Hum. Evol. **48**, 85–96.
Riede, T., Bronson, E., Hatzikirou, H., and Zuberbühler, K. (**2006**). "Multiple discontinuities in nonhuman vocal tracts—A response to Lieberman (2006)," J. Hum. Evol. **50**, 222–225.
Riede, T., Suthers, R. A., Fletcher, N., and Blevins, W. (**2006**). "Songbirds tune their vocal tract to the fundamental frequency of their song," Proc. Natl. Acad. Sci. U.S.A. **103**, 5543–5548.
Riede, T., Wilden, I., and Tembrock, G. (**1997**). "Subharmonics, biphonations, and frequency jumps-common components of mammalian vocalization or indicators for disorders?" Z. Säugetierkunde (Suppl. II) **62**, 198–203.
Scherer, R. C., Shinwari, D., DeWitt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle," J. Acoust. Soc. Am. **109**, 1616–1630.
Schneider, R. (**1964**). "Der Larynx der Säugetiere (The larynx of mammals)," Handbuch der Zoologie **5**, 1–128.
Schön-Ybarra, M. (**1995**). "A comparative approach to the non-human primate vocal tract: Implications for sound production," in *Current Topics in Primate Vocal Communication*, edited by E. Zimmermann, J. D. Newman, and U. Juergens (Plenum, New York), pp. 185–198.
Sohndi, M. M., and Schroeter, J. (**1987**). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Acoust., Speech, Signal Process. **35**, 955–967.
Starck, D., and Schneider, R. (**1960**). "Respirationsorgane, Larynx," Primatologia **3**, 423–587.
Tembrock, G. (**1974**). "Sound production in *Hylobathes and Symphalangus*," in *Gibbon and Siamang*, edited by D. Rumbaugh (Karger, Basel), Vol. **3**, pp. 176–205.
Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.
Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R., and Story, B. H. (**1997**). "Acoustic interaction of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.
Titze, I. R., Schmidt, S. S., and Titze, M. R. (**1995**). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **97**, 3080–3084.
Titze, I. R. (**2004a**). "A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice," J. Voice **18**, 292–298.
Titze, I. R. (**2004b**). "Theory of glottal airflow and source-filter interaction in speaking and singing," Acta Acust. **90**, 641–648.
Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.
Titze, I. R., Riede, T., and Popollo, P. (**2008**). "Nonlinear source-filter coupling in phonation: Vocal exercises," J. Acoust. Soc. Am. **123**, 1902–1915.
Whitehead, J. M. (**1995**). "Vox Alouattinae: A preliminary survey of the acoustic characteristics of long-distance calls of howling monkeys," Int J. Primatol. **16**, 121–144.
Zhang, Z., Neubauer, J., and Berry, D. A. (**2006a**). "The influence of subglottal acoustics on laboratory models of phonation," J. Acoust. Soc. Am. **120**, 1558–1569.
Zhang, Z., Neubauer, J., and Berry, D. A. (**2006b**). "Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds," J. Acoust. Soc. Am. **120**, 2841–2849.