RESEARCH ARTICLE

Japanese macaque phonatory physiology

Christian T. Herbst[1,*], Hiroki Koda[2], Takumi Kunieda[2], Juri Suzuki[2], Maxime Garcia[1,3], W. Tecumseh Fitch[1] and Takeshi Nishimura[2]

ABSTRACT

Although the call repertoire and its communicative function are relatively well explored in Japanese macaques (*Macaca fuscata*), little empirical data are available on the physics and the physiology of this species' vocal production mechanism. Here, a 6 year old female Japanese macaque was trained to phonate under an operant conditioning paradigm. The resulting 'coo' calls and spontaneously uttered 'growl' and 'chirp' calls were recorded with sound pressure level (SPL) calibrated microphones and electroglottography (EGG), a non-invasive method for assessing the dynamics of phonation. A total of 448 calls were recorded, complemented by *ex vivo* recordings on an excised Japanese macaque larynx. In this novel multidimensional investigative paradigm, *in vivo* and *ex vivo* data were matched via comparable EGG waveforms. Subsequent analysis suggests that the vocal range (range of fundamental frequency and SPL) of the macaque was comparable to that of a 7–10 year old human, with the exception of low intensity chirps, the production of which may be facilitated by the species' vocal membranes. In coo calls, redundant control of fundamental frequency in relation to SPL was also comparable to that in humans. EGG data revealed that growls, coos and chirps were produced by distinct laryngeal vibratory mechanisms. EGG further suggested changes in the degree of vocal fold adduction *in vivo*, resulting in spectral variation within the emitted coo calls, ranging from 'breathy' (including aerodynamic noise components) to 'non-breathy'. This is again analogous to humans, corroborating the notion that phonation in humans and non-human primates is based on universal physical and physiological principles.

KEY WORDS: Voice production principles, Laryngeal configuration, Electroglottography, Primate, Voice range profile, Excised larynx preparation

INTRODUCTION

Humans and non-human primates (together with other mammals) are believed to share a universal mechanism of phonation (laryngeal sound production), governed by the myoelastic aerodynamic (MEAD) principle (van den Berg, 1958; Titze, 2006; Herbst, 2016). Steady airflow, coming from the lungs, is converted into a sequence of airflow pulses by the passively vibrating vocal folds (and/or other laryngeal tissues), resulting in self-sustaining oscillation. The acoustic pressure waveform generated by this sequence of flow pulses excites the vocal tract, which filters them acoustically, and the result is radiated from the mouth (and/or the nose) (Story, 2002). This phenomenon, combining the individual contributions of the laryngeal sound source and the vocal tract to determine the quality of the emitted sound, is termed the source–filter theory of sound production (Fant, 1960; Chiba and Kajiyama, 1941; Taylor et al., 2016; Fitch and Hauser, 1995) and its non-linear extension (Titze, 2008; Flanagan, 1968; Rothenberg, 1981).

In human speech and singing, the physics and physiology of phonation and the respective detailed motor control are relatively well investigated, owing to several decades of research *in vivo* (Baken and Orlikoff, 2000), *ex vivo* (Döllinger et al., 2011) and *in silico* (Kob, 2003; Story, 2002). In contrast, much less is known about the actual physical and functional and/or physiological framework of *in vivo* sound production in non-human mammals. The non-human vocal system is typically treated as a 'black box', and its function is inferred from the acoustic output alone. This is true, for instance, for the vocalization of Japanese macaques (*Macaca fuscata* Blyth 1875). Ever since Itani's groundbreaking work (Itani, 1963), the investigation of this species' vocal communication has received widespread attention (Le Prell and Moody, 1997; Beecher et al., 2008; Blount, 1985; Katsu et al., 2016; Green, 2010; Tokuda et al., 2002; Machida, 1990; Masataka, 2010; Owren et al., 1992; Sugiura, 2008; Bouchet et al., 2017; Koda, 2004). However, most studies have typically focused on the acoustic description and classification of calls, to be regarded in a motivational and social context.

The purpose of this study is thus to provide physiological evidence concerning laryngeal *in vivo* sound production in Japanese macaques. Addressing the hypothesis that humans and non-human primates share universal sound production principles, the gathered data were compared with that of humans, in order to demonstrate detailed functional similarities.

The compliance of humans with measurement protocols allows for *in vivo* documentation of a number of physical and physiological key variables of human speech production and singing, such as subglottal and/or tracheal air pressure (Schutte, 1980; Finnegan et al., 1998), glottal airflow (Rothenberg, 1977; Stathopoulos and Weismer, 1985), laryngeal configuration (Herbst et al., 2011; Södersten et al., 1995), vocal tract geometry (Echternach et al., 2008; Story et al., 2003) or the kinematics of vocal fold vibration (Hertegard, 2005; Deliyski and Hillman, 2010; Lohscheller and Eysholdt, 2008). Unfortunately, most of the involved investigative methods are somewhat uncomfortable or invasive, which makes application to non-human primates a challenge.

A non-invasive alternative for assessing the dynamics of laryngeal vocal fold vibration during sound production is electroglottography (EGG) (Baken, 1992; Fabre, 1957). A high frequency, low intensity current is passed between two electrodes attached to either side of the skin at the side of the thyroid cartilage at the level of the vocal folds (see Fig. 1A). The measured admittance variations are largely proportional to the time-varying

[1]Bioacoustics Laboratory, Department of Cognitive Biology, University Vienna, Althanstrasse 14, 1090 Vienna, Austria. [2]Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan. [3]ENES Lab, Université Lyon/Saint-Etienne, NEURO-PSI, CNRS UMR 9197, 23 rue Paul Michelon, 42023 Saint-Etienne, France.

*Author for correspondence (herbst@ccrma.stanford.edu)

C.T.H., 0000-0001-9095-3953; M.G., 0000-0003-2014-7387

Received 11 October 2017; Accepted 26 March 2018

vocal fold contact area (Hampala et al., 2016), thus providing detailed physiological information on vocal fold vibration. A schematic model of a stereotypical EGG signal for one vibratory cycle of the vocal folds in humans is shown in Fig. 1B (Berke et al., 1987; Baken and Orlikoff, 2000). The landmarks in that illustration are identified as follows: (a) initial contact of the lower vocal fold margins; (b) initial contact of the upper vocal fold margins; (c) maximum vocal fold contact reached (glottis not necessarily fully closed); (d) de-contacting phase by separation of the lower vocal fold margins; (e) upper margins start to separate; and (f) glottis is open and the contact area is at its minimum.

Several approaches exist for extracting quantitative information from the raw EGG signal (Rothenberg and Mahshie, 1988; Orlikoff, 1991; Baken and Orlikoff, 2000). These are loosely correlated with physical key phenomena of vocal fold vibration, but need to be interpreted with care (Herbst et al., 2017, 2014).

Although EGG, thanks to its relatively inexpensive and non-invasive nature, has seen wide application in human voice science, surprisingly, only one pilot study has been conducted on non-human primates (Brown and Cannito, 1995). Here, we apply EGG data acquisition to *in vivo* phonation of a female Japanese macaque trained to vocalize on command. EGG data are complemented with sound pressure level (SPL) calibrated acoustic recordings and matched EGG data from an excised larynx preparation of a Japanese macaque *ex vivo*. This novel multidimensional approach allows for deeper insights into the physiological and physical nature of voice production in this species.

MATERIALS AND METHODS

Data acquisition *in vivo*

In vivo data acquisition was performed at the Primate Research Institute, Inuyama, Aichi, Japan. All procedures were approved by the ethics committee of the Primate Research Institute of Kyoto University (number 2015-014, 2016-103), with compliance to the Guide for the Care and Use of Laboratory Primates (third edition, the Primate Research Institute, Kyoto University, 2010). The subject animal was a 6.5 year old female Japanese macaque, with a resting vocal fold length of approximately 7.7 mm, as measured from a computerized tomography (CT) scan with a spatial resolution of 0.35×0.35 mm and a slice interval of 0.2 mm.

The animal had been trained over a period of 6 months for another research project (H.K., T.K. and T.N., unpublished) to sit in a custom-made monkey chair wearing a special-purpose jacket (Fig. 1A). Using an operant conditioning approach, the animal was rewarded when producing ‘coo’ calls after presentation of a visual and auditory stimulus. In addition to these trained responses, we also recorded a number of spontaneous calls (see below). For the purpose of this work, a total of three recording sessions, each lasting approximately 50 min, were conducted over a period of 8 days.

EGG signals were recorded with a VoceVista electroglottograph (Roden, The Netherlands). The EGG electrodes were embedded into the collar of a special-purpose jacket that was worn by the animal during data acquisition (see Fig. 1A). In this setup, head movement of the animal resulted in intermittent contact loss between the electrodes and the individual’s neck in approximately 60% of all recorded signals. EGG signals were only considered for further analysis if two conditions were fulfilled: (1) a cyclical EGG signal at a fundamental frequency (f_o) corresponding to that of the acoustic signal (checked through inspection of respective spectrograms) was present; and (2) there was no evidence of clipping in the acquired EGG signal.

The acoustic signal was recorded with a Sennheiser MKE platinum-C microphone (Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany). The microphone was placed at a fixed distance of 10 cm from the animal’s mouth. SPL levels were calibrated with C frequency weighting for a distance of 30 cm using an SPL meter (ATL SL-8851, ATP Instrumentation Ltd, Ashby-de-la-Zouch, UK), applying method 5 from Svec and Granqvist (2017). Background noise levels were measured at 55.3 dB(C).

Both the EGG and the acoustic signal were simultaneously digitized at a sampling frequency of 48 kHz with a Tascam audio interface (US-144KMII, TEAC America Inc., Montebello, CA, USA). The digitized signals were recorded using the software Audacity (http://www.audacityteam.org/) and stored as 16-bit uncompressed stereo .wav files.

Data acquisition *ex vivo*

Data acquisition *ex vivo* was conducted at the Department of Cognitive Biology, University of Vienna, Austria. No ethical approval was required. The larynx was from a female Japanese macaque (mass=7.4 kg, head–body length without tail=72.6 cm) that had died of natural causes, acquired through the specimen acquisition program at the National Museums of Scotland. A detailed description of that specimen’s preparation is provided

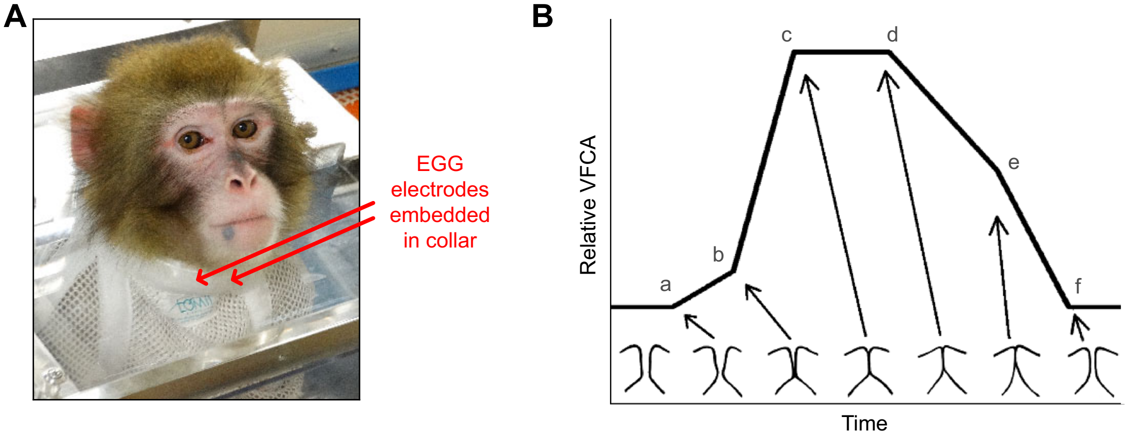

Fig. 1. Electroglottography (EGG). (A) Attachment of EGG electrodes in the experimental setup. (B) Schematic diagram of a stereotypical human EGG waveform for one glottal cycle (Baken and Orlikoff, 2000; Berke et al., 1987), with illustrations of vocal fold movement and contact (a to f) within the coronal plane shown at the bottom (see Introduction). VFCA, vocal fold contact area.

Table 1. Mean±s.d. fundamental frequency (f_o), dominant frequency (f_{DOM}), sound pressure level (SPL) and harmonics-to-noise ratio (HNR) for all call types

Call type	No. of calls	No. of data points	Mean f_o (Hz)	Mean f_{DOM} (Hz)	Mean SPL [dB(C)]	Mean HNR (dB)
Growl	31	3571	296.1±142.5	488.2±279.7	75.8±5.4	6.4±4.8
Coo	377	127,981	585.0±74.1	725.7±319.5	78.3±7.0	22.7±7.0
Chirp	14	747	3134.0±559.2	3127.7±702.6	77.5±11.8	3.1±2.5

elsewhere (Garcia et al., 2017). The resting vocal fold length was visually determined to be approximately 7.3 mm.

A previously described excised larynx setup was utilized (Herbst et al., 2014). The larynx was mounted on a vertical tube supplying heated (ca. 37°C) and humidified (100% humidity) air. For the purpose of this study, the vocal folds were adducted and elongated manually, in order to have maximum freedom for achieving vocalizations that resemble those documented *in vivo*.

Vocal fold vibration was documented with acoustic and EGG recordings (see Herbst et al., 2014 for details), while simultaneously measuring the subglottal driving (air) pressure. For comparative analysis of data recorded *in vivo* and *ex vivo*, EGG signals from these two scenarios were matched using the following criteria: (1) comparable f_o; (2) comparable periodicity and harmonic content (nearly periodic and sinusoidal for coo calls, slightly irregular and slightly aperiodic for growls and chirps); and (3) comparable relative EGG signal level (note that the EGG signal level of chirp calls was typically approximately 15–20 dB lower than that of all other calls; see below).

Data analysis

f_o was estimated with the Praat (Boersma and Weenink, 2017) program's autocorrelation-based algorithm ['To Pitch (ac)…']. Standard parameters were used, except for minimum and maximum f_o, which were set to 50 and 5000 Hz, respectively. f_o was estimated every millisecond, resulting in 1000 analysis data points per second.

At the time offset of each successfully estimated f_o data point, two further parameters were calculated with a custom algorithm written in Python by C.T.H.: the calibrated SPL, expressed in dB(C), and the dominant frequency (f_{DOM}) (Fischer et al., 2013), representing the frequency with the maximum amplitude within the acoustic spectrum of the analyzed signal portion. The respective source code is available online (www.christian-herbst.org/python/).

Preliminary perceptual assessment of the acoustic data suggested various degrees of 'breathiness' (i.e. aerodynamic noise components) in a subset of the coo calls produced *in vivo*. In order to assess this quantitatively, the average harmonics-to-noise ratio (HNR) was calculated for all coo calls with Praat. In particular, the function 'To Harmonicity (ac)' was called with standard parameters, except for the time step (1 ms) and minimum f_o (50 Hz).

Glottal efficiency (E_{GL}) is a measure of aerodynamic energy conversion during sound production. It is the ratio of radiated acoustic power (i.e. the system's output) to aerodynamic power (i.e. the system's input) (van den Berg, 1956; Bouhuys et al., 1968; Schutte, 1980). Glottal efficiency, expressed in dB, was determined here as:

$$E_{GL} = 10\log_{10}\frac{P_{RAD}}{P_{AIR}}, \quad (1)$$

where P_{RAD} is the radiated power and P_{AIR} is the aerodynamic power. P_{RAD} was calculated in W as:

$$P_{RAD} = 4\pi r^2 I, \quad (2)$$

where r is the microphone distance (30 cm in this case) and I is the sound intensity in W m^{-2}, derived from the measured SPL at 30 cm as:

$$I = I_0 10^{SPL/10}, \quad (3)$$

where the reference sound intensity I_0=10^{-12} W m^{-2}. Finally, the aerodynamic power, P_{AIR}, expressed in W, was calculated as the product of the time-averaged glottal airflow and the time-averaged subglottal pressure.

RESULTS

A total of 448 calls were recorded *in vivo*, which were labeled manually according to the classification scheme provided by Green (1975), resulting in 377 coos, 31 growls, 14 chirps, and 26 transitions between coo and grunt. Whereas the coo calls were emitted as a trained response of the investigated animal, the growls and chirps were mostly spontaneous vocal emissions uttered when one of the experimenters adjusted the EGG electrodes.

An overview of analysis data for all calls is provided in Table 1. The relationship between f_o and SPL for all vocalizations is depicted in Fig. 2A. Such a display, called a phonetogram (Damste, 1970) or voice range profile (VRP) (Pabon and Plomp, 1988), is a typical tool in human voice science and clinical work, utilized to obtain an overview of a person's vocal capacities. The gray diamonds and dashed lines superimposed upon Fig. 2A, allowing for a comparison between the investigated Japanese macaque and humans, are normative VRP data for children aged 7–10 years (Schneider et al., 2010).

In order to corroborate the similitude of VRP data between Japanese macaques and human children on an anatomical level, the vocal fold lengths of the Japanese macaques analyzed *in vivo* and *ex vivo* (7.7 and 7.3 mm, respectively) were compared with those of pre-pubertal children according to data from Hirano et al. (1983) (Fig. 2B). A substitution of the vocal fold lengths of the two examined Japanese macaque specimens into the linear regression through the data for children below 12 years of age (Hirano et al., 1983) suggests that comparable vocal fold lengths are found in children aged approximately 7.9 and 7.4 years, respectively.

Preliminary analysis of the coo calls suggested a systematic co-variation between f_o and SPL in a large portion of the calls (see Fig. 2C for an example). This co-variation was quantified by calculating first order linear regressions between SPL and f_o within all coo calls. Computing the average of all data points where the coefficient of determination, R^2, was equal to or greater than 0.8

Table 2. Subglottal pressure (P_{SUB}), airflow, SPL and glottal efficiency (E_{GL}) for the three excised larynx phonations depicted in Fig. 3

Call type	P_{SUB} (kPa)	Airflow (l min^{-1})	SPL [dB(C) at 30 cm]	E_{GL} (dB)
Growl	3.4	34.7	81.0	−41.3
Coo	2.4	55.1	74.6	−48.3
Chirp	4.7	37.5	89.9	−34.2

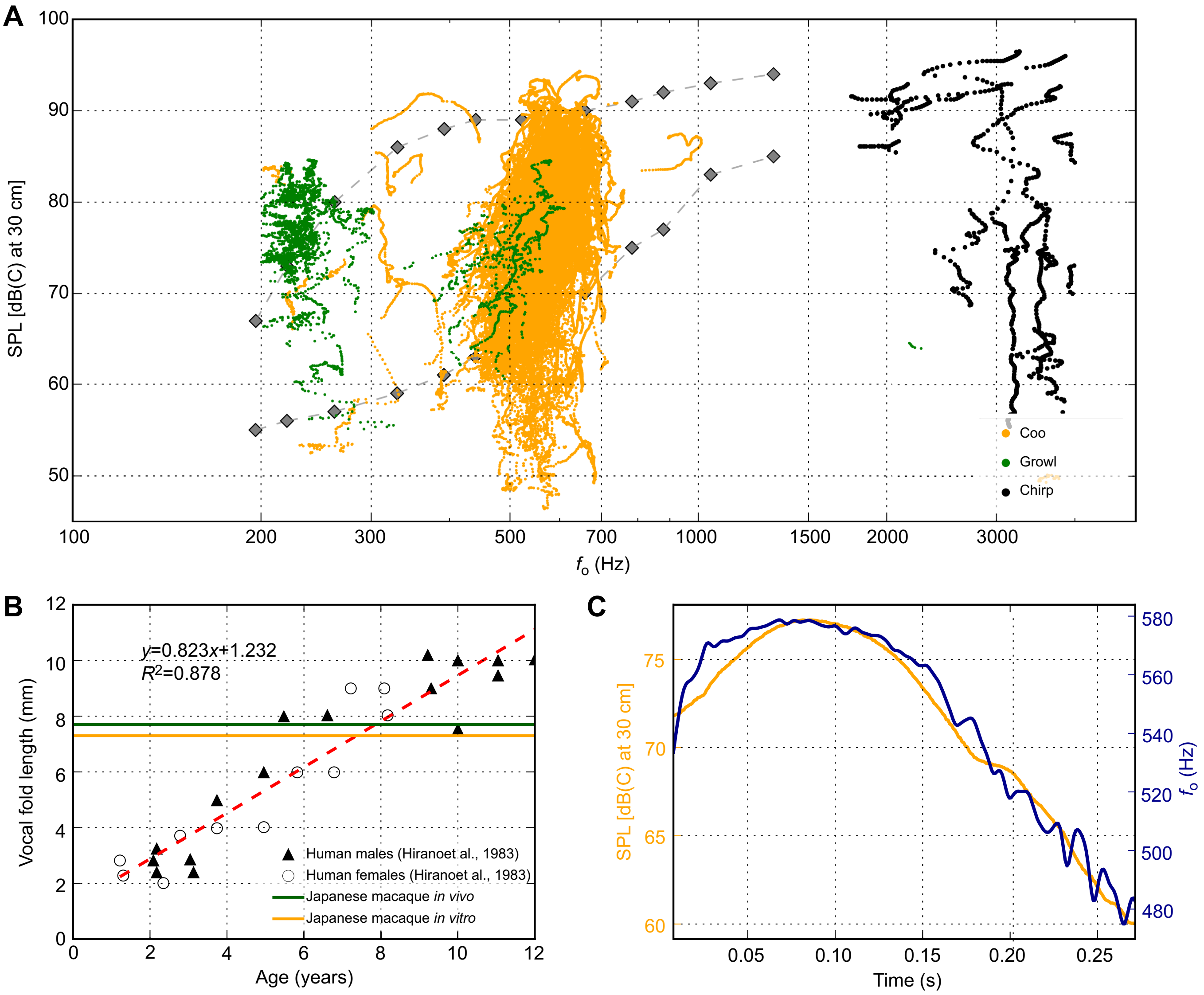

Fig. 2. Fundamental frequency (f_o) and sound pressure level (SPL) of analyzed vocalizations. (A) Phonetogram showing SPL versus f_o. The superimposed diamonds and dashed lines represent normative voice range profile (VRP) data from human children aged 7–10 years (Schneider et al., 2010). (B) Vocal fold length measurements for pre-pubertal children (Hirano et al., 1983), with superimposed vocal fold length measurements from the two Japanese macaques investigated *in vivo* and *ex vivo*. (C) SPL and f_o contour for a selected coo call.

(39.3% of all cases) resulted in an average slope of 0.28 semitones per dB SPL. The semitone scale (Young, 1939) was chosen in order for the data to be comparable with those in a previous publication in humans (Gramming et al., 1988). For reference purposes, at the mean f_o of all coo calls, this value would be equivalent to an increase of approximately 9.5 Hz per dB SPL.

Basic physical data for the excised larynx sound production are listed in Table 2: subglottal pressure, airflow rates, SPL and glottal efficiency. In Fig. 3, stereotypical EGG waveforms from both the *in vivo* condition and the excised larynx preparation are shown for all three call types. Care has been taken to find EGG waveforms that are similar both in appearance and in f_o. The EGG waveforms for the growl vocalizations were mostly irregular, with residual traces of periodicity. The coo calls typically resulted in periodic EGG waveforms, approximating a sinusoidal shape in most cases (but see Fig. 5 for an important counter-example). The EGG signals of the chirps also approximated sinusoidal shapes. However, they had markedly weaker amplitudes (−26.6 dB in Fig. 3, compared with −8 dB and −11 dB for growls and coos, respectively). This suggests a lesser degree of vocal fold contact, and noise introduced by the measurement equipment had greater influence on the waveform.

In 26 out of the 448 analyzed calls, transitions between the coo and growl call types were found. These transitions typically occurred over a few glottal vibratory cycles. One such example is documented in Fig. 4: f_o drops abruptly from approximately 464 Hz to approximately 190 Hz, whereas the EGG waveform abruptly alternates between two distinct shapes around t=280 ms in Fig. 4D.

The average HNR of all coo calls is plotted against the respective average SPL in Fig. 5. The data in panel A suggest an overall trend for HNR to be lower in softer calls. A stereotypical example of a coo call characterized as 'breathy' (including aerodynamic noise components) by the experimenters is further analyzed in panels B and C. The spectrogram of the acoustic signal contained only three harmonics above noise level, and the respective EGG waveform was quasi-sinusoidal, containing considerable noise. In contrast, the

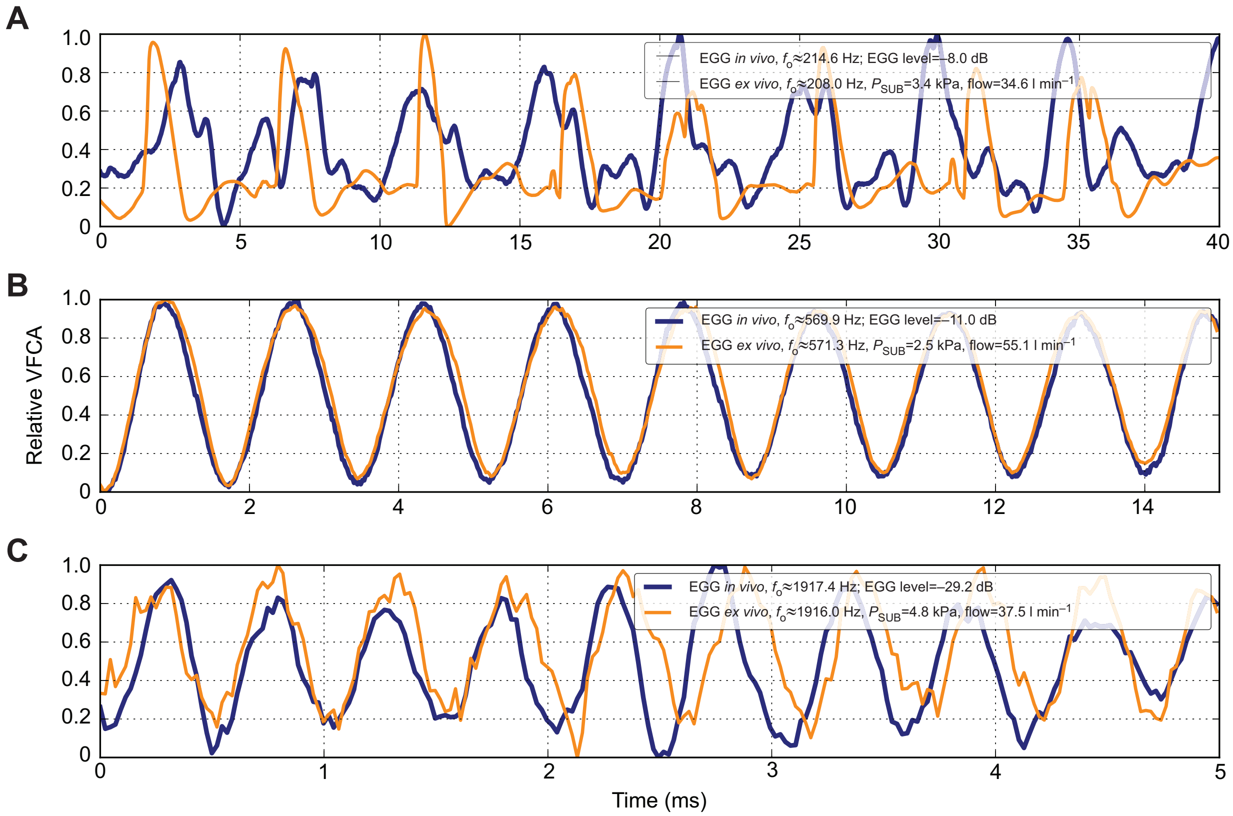

Fig. 3. Comparable EGG waveforms of *in vivo* and excised larynx recordings for all three call types. (A) Growl (issuing a threat). (B) Coo (contact call). (C) Chirp (defensive withdrawal).

acoustic signal of a stereotypical coo call characterized as ‘non-breathy’ (panels D and E) contained 12 harmonics above the noise floor, and the corresponding EGG waveform was devoid of visible noise components, resulting in a pronounced wave shape.

DISCUSSION

This study introduces a new multidimensional investigative paradigm to the fields of primatology and animal bioacoustics: controlled *in vivo* experiments with accompanying excised larynx experimentation, linked through matched EGG waveforms as a physiological ‘ground truth’. In this manner, advantages from both approaches can be combined. The *in vivo* setup, thanks to calibrated microphone signals and a controlled mouth-to-microphone position, facilitates assessment of SPLs of targeted call types (see Fig. 2). The supplemented data from the excised larynx experiment allow for the estimation of physical and physiological voice production parameters (see Table 2), which are difficult to obtain *in vivo*. In this approach, EGG data provide the key evidence through which the two setups (*in vivo* versus excised larynx) are linked. Although in the current study, larynges of two different animals were examined *in vivo* and *ex vivo*, future investigations could, given logistical and ethical feasibility, utilize the same animal in both setups to control for variation in laryngeal anatomy between animals.

The three investigated call types, growls, coos and chirps, had distinct fundamental frequencies and were well separated within the generated phonetogram (Fig. 2). The growl and coo calls were well aligned within normative VRP data published for 7–10 year old children (Schneider et al., 2010) (but note the greater sound levels of the growl vocalizations in comparison with the respective phonations of children around 200–250 Hz). However, even the higher frequencies of the chirps (f_o≈3 kHz) can be sung by some children of that age, but typically only at high vocal intensities (C.T.H., personal observation). The VRP comparison is, however, limited by the fact that the VRP data of the children were acquired via instructions to continuously and fully cover their entire voice range (i.e. reaching the minima and maxima of both sound level and f_o), whereas the data from the Japanese macaque were acquired through the operant conditioning approach without such restrictions. The actual voice range of the Japanese macaque could thus be greater than that indicated by the collected data. Furthermore, although the children’s VRP is continuous, the Japanese macaque’s VRP is not, owing to the different methods of data acquisition. Therefore, it cannot be determined whether areas in the Japanese macaque’s VRP that are not covered by our current data from growls, coos or chirps (e.g. the frequency region between 750 and 1700 Hz) constitute evidence that the animal would not have the ability to produce sounds at those frequencies and sound levels.

In humans, the f_o of vocal fold vibration can be approximated with a simple string model as:

$$f_o = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}}, \tag{4}$$

where L is the vocal fold length, σ is the stress within the vocal fold and ρ is the tissue density (assumed to be constant) (Titze, 2000). Although the stress (and hence the vocal fold elongation) can be varied individually (Titze et al., 2016), the resting (i.e. unstretched) vocal fold length can be assumed to be constant for an individual.

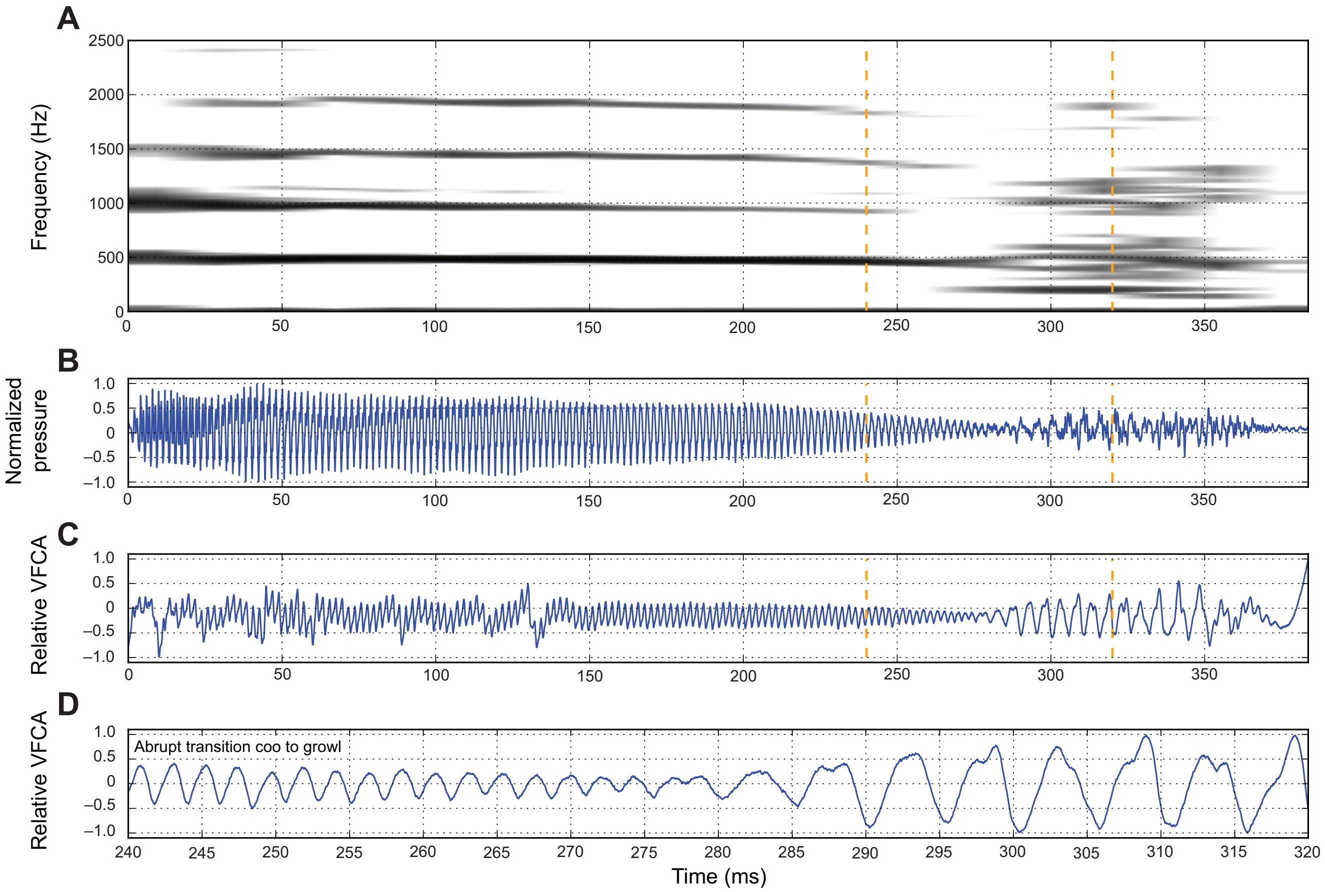

Fig. 4. Abrupt transition from coo to lower frequency growl. (A) Narrow-band spectrogram of microphone signal. (B) Acoustic signal. (C) EGG signal. (D) Portion of the EGG signal, extracted at *t*=240–320 ms.

Vocal fold length is thus a main anatomical determinant for an individual's f_o range.

A recent comparative allometric study showed that vocal fold length is a good predictor for the minimum f_o across 11 non-human primate species (Garcia et al., 2017). The resting vocal fold length of the Japanese macaques investigated here *in vivo* and *ex vivo* was approximately 7.7 and 7.3 mm, respectively. Hirano et al. (1983) found comparable vocal fold lengths for children aged approximately 6–10 years (see Fig. 2B). This evidence thus strongly suggests that the similar f_o ranges of the examined Japanese macaque and 7–10 year old children are determined by comparable vocal fold length. This would imply that the string model approximation (Eqn 4) applies to both humans and non-human primates (see also Riede, 2010), supporting the hypothesis of universal sound production principles.

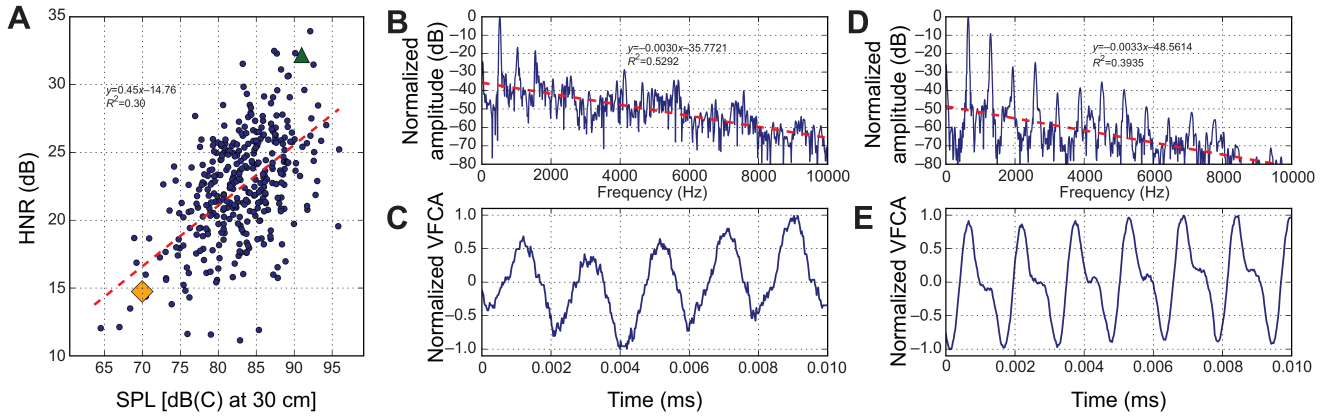

Fig. 5. Variation in spectral quality between coo calls. (A) Average harmonics-to-noise ratio (HNR) as a function of average SPL per call (one blue circle represents one coo call, *n*=357). The calls identified with an orange diamond and a green triangle are described in more detail in panels B–E. (B) Acoustic frequency spectrum and (C) EGG waveform for a stereotypical breathy case – see orange diamond in panel A. (D) Acoustic frequency spectrum and (E) EGG waveform for a stereotypical non-breathy case – see green triangle in panel A.

The similarity between the primate and the human vocal organ is also seen when assessing dynamical aspects of f_o control. We found an f_o increase of approximately 0.28 semitones per dB SPL. This value is comparable to data from humans, where an increase of approximately 0.4 semitones per dB SPL was found (Gramming et al., 1988). In analogy to the argument made in that study (Gramming et al., 1988) and building on previous research in humans, we hypothesize that subglottal pressure (van den Berg and Tan, 1959; Titze, 1989) is a major influencing factor for f_o control in Japanese macaque vocalizations (the other being vocal fold tension; Titze et al., 2016), thus further demonstrating the physiological commonality between Japanese macaques and humans. Rigorous testing of that hypothesis with excised larynx experiments is, however, required.

Normative VRP data from humans suggest that the upper f_o limit can typically only be reached at maximum SPLs (Sulter et al., 1994), suggesting high subglottal pressures (Schutte, 1980). In contrast, the investigated Japanese macaque's chirp vocalizations *in vivo* were produced at relatively low SPLs, a phenomenon which deserves further discussion. We hypothesize that these low SPLs were facilitated by the presence of vocal membranes (sometimes called 'vocal lips') in the laryngeal anatomy of the Japanese macaque, i.e. thin upward extensions of the vocal folds with little mass (Fitch, 2002; Schön Ybarra, 1995; Mergell et al., 1999). Unfortunately, we were unable to duplicate these softer chirp vocalizations in the one specimen examined in the excised larynx setup. Further investigation with excised larynx experiments and computational modeling is thus necessary to substantiate this hypothesis.

Exemplary EGG evidence suggested distinct differences in vocal fold vibration patterns for the three call types. The sinusoidal waveforms of the coo calls in Figs 3 and 5C, as well as the chirp call in Fig. 3, are comparable to EGG data from humans phonating in the so-called falsetto register (thyroarytenoid muscle not contracted) with a low degree of vocal fold adduction (Herbst et al., 2017), regularly resulting in a posterior glottal gap and breathy phonation (Sundberg, 1995). This class of EGG signals typically has a low signal amplitude owing to the lack of vocal fold contact.

Interpretation of the other EGG waveforms, including those presented in Figs 3A and 5E, is more difficult because they do not clearly match stereotypical waveforms known from humans. This can be attributed to potential differences in laryngeal anatomy between humans and Japanese macaques. In EGG, the complex three-dimensional (de-)contacting pattern of the vocal folds is reduced to a one-dimensional value, reflecting the time-varying relative vocal fold contact area. Consequently, anatomically induced differences of vocal fold geometry are reflected in the resulting EGG waveform. Further excised larynx experiments with acquisition of simultaneous EGG and high-speed video data are thus necessary to better facilitate interpretation of EGG waveforms in Japanese macaques and other primate species.

This limitation notwithstanding, EGG was quite useful for revealing the dynamics of laryngeal sound generation *in vivo*. This is perhaps best seen in Fig. 4, where a transition from coo to growl is documented. The EGG evidence reveals an abrupt transition between two distinct states of vocal fold vibration, occurring over the course of about five vibratory cycles. Several insights can be gained from this example: (1) the cause for acoustic differences between these call types is clearly laryngeal, similar to different laryngeal mechanisms in human voice registers (Henrich, 2006); (2) the suddenness of the change between the two call types is evidence for the presence of a bifurcation, i.e. an abrupt change between vibratory states of a non-linear system when gradually varying boundary conditions (Fitch et al., 2002; Herzel et al., 1998); and (3) as expected from a bifurcating system, the two vibratory phenomena do not coexist.

Some of the softer coo calls had a pronounced breathy perceptual quality, as noticed by the experimenters. This phenomenon, which is spectrally characterized by fewer noteworthy harmonics and the appearance of high-frequency noise components, was quantified by calculation of the HNR (see Fig. 5A). Acoustically, the coo calls with lower HNR (see Fig. 5B,C for a stereotypical example) typically had only about two to five harmonics above the noise floor. The respective EGG signals assumed a sinusoidal wave shape, with superimposed noise. As mentioned above, this is analogous to breathy phonation in the falsetto register in humans (Herbst et al., 2017) and strongly suggests that those breathy coo vocalizations were produced with incomplete glottal closure, allowing turbulent airflow to occur, thus causing the audible noise components and giving the perceptual impression of breathiness.

The breathy coo vocalizations were contrasted by non-breathy coo vocalizations, which typically had higher HNR values. The corresponding EGG waveforms were less noisy, deviated from a sinusoidal shape, and bore indicators of vocal fold contacting and de-contacting events, suggesting a greater degree of vocal fold adduction than in the breathy calls. However, as mentioned above, without clearly established landmarks for EGG signals in Japanese macaques, further interpretation requires caution.

Overall, the physiologically based EGG evidence strongly suggests that the investigated macaque varied its glottal configuration while producing the variety of coo calls *in vivo*. This is, to our knowledge, a novel finding that has not yet been documented at the laryngeal level for vocalizations in non-human primates and other mammals. Laryngeal modification of the voice timbre (i.e. the spectral composition of the sound source) via the degree of glottal adduction would provide an animal with an additional dimension for voice quality modification, potentially allowing macaques to encode arousal and/or valence states in a social communicative context, analogous to what has been shown for humans when using breathy voice in speech (Gobl and Ní Chasaide, 2003; Ishi et al., 2010; Miyazawa et al., 2017).

This study has a few limitations that are worth mentioning. This is a two-subject study, so findings may not be generalized without further evidence. The larynx utilized for the *ex vivo* experiments was not flash-frozen post mortem (Chan and Titze, 2003), which might have altered the biomechanical tissue properties, thus explaining some surprisingly high values for subglottal pressure and airflow (see Table 2). Repetition of the experiments with flash-frozen larynx specimens is thus warranted.

Conclusions

A novel multidimensional investigative paradigm was introduced with this study: controlled *in vivo* data acquisition, supplemented by *ex vivo* recordings from an excised larynx setup, linked via matched EGG waveforms. The data from these experiments, although from only two animals, provide a number of new insights into the sound production of Japanese macaques. When considering growls, coos and chirps, the vocal range of the investigated adult Japanese macaque was comparable to that of a 7–10 year old human, with the exception of low intensity chirps, the production of which may be facilitated by the species' vocal membranes. In coo calls, dynamic control of f_o in relation to SPL was also comparable to that in humans. EGG evidence suggested that growls, coos and chirps were produced by distinct laryngeal vibratory mechanisms, analogous to those of humans. EGG data also revealed that the investigated

Japanese macaque most likely varied the degree of vocal fold adduction, resulting in variations of the spectral characteristics within the emitted coo calls, ranging from breathy to non-breathy. This is again analogous to what is found in humans, further corroborating the hypothesis that humans and non-human primates share universal physical and physiological principles of vocal production, governed by the MEAD principle.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: C.T.H., H.K., T.N.; Methodology: C.T.H., H.K., T.K., M.G., T.N.; Software: C.T.H., H.K.; Validation: C.T.H.; Formal analysis: C.T.H.; Investigation: C.T.H., H.K., T.K., J.S., M.G., T.N.; Resources: C.T.H., H.K., J.S., W.T.F., T.N.; Writing - original draft: C.T.H.; Writing - review & editing: C.T.H., H.K., M.G., W.T.F., T.N.; Visualization: C.T.H.; Supervision: C.T.H., W.T.F., T.N.; Project administration: T.N.; Funding acquisition: T.N.

Funding

This research was supported by an APART grant from the Austrian Academy of Sciences (awarded to C.T.H.), a postdoctoral fellowship from the Fyssen Foundation (awarded to M.G.), and the Japan Society for the Promotion of Science KAKENHI grants no. 16H04848 (awarded to T.N.) and no. 4903, JP17H06380 (awarded to H.K.).

References

Baken, R. J. (1992). Electroglottography. *J. Voice* **6**, 98-110.

Baken, R. J. and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*, 2nd edn. San Diego, CA: Singular Publishing, Thompson Learning.

Beecher, M. D., Petersen, M. R., Zoloth, S. R., Moody, D. B. and Stebbins, W. C. (2008). Perception of conspecific vocalizations by Japanese macaques. *Brain Behav. Evol.* **16**, 443-460.

Berke, G., Moore, D. M., Hanson, D. G., Hantke, D. R., Gerratt, B. R. and Burstein, F. (1987). Laryngeal modeling: theoretical, in vitro, in vivo. *Laryngoscope* **97**, 871-881.

Blount, B. G. (1985). "Girney" vocalizations among Japanese macaque females: context and function. *Primates* **26**, 424-435.

Boersma, P. and Weenink, D. (2017). Praat: doing phonetics by computer. Available at: http://www.praat.org.

Bouchet, H., Koda, H. and Lemasson, A. (2017). Age-dependent change in attention paid to vocal exchange rules in Japanese macaques. *Anim. Behav.* **129**, 81-92.

Bouhuys, A. Mead, J., Proctor, D. F. and Stevens, K. N. (1968). Pressure-flow events during singing. *Ann. N. Y. Acad. Sci.* **155**, 165-176.

Brown, C. H. and Cannito, M. P. (1995). Modes of vocal variation in Syke's monkey (Cercopithecus albogularis) squeals. *J. Comp. Psychol.* **109**, 398-415.

Chan, R. W. and Titze, I. R. (2003). Effect of postmortem changes and freezing on the viscoelastic properties of vocal fold tissues. *Ann. Biomed. Eng.* **31**, 482-491.

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

Damste, P. H. (1970). The phonetogram. *Pract Otorhinolaryngol (Basel)* **32**, 185-187.

Deliyski, D. D. and Hillman, R. E. (2010). State of the art laryngeal imaging: research and clinical implications. *Curr. Opin Otolaryngol. Head Neck Surg.* **18**, 147-152.

Döllinger, M., Kobler, J., Berry, D. A., Mehta, D. D., Luegmair, G. and Bohr, C. (2011). Experiments on analysing voice production: excised (human, animal) and in vivo (animal) approaches. *Curr. Bioinform.* **6**, 286-304.

Echternach, M. Sundberg, J., Arndt, S., Breyer, T., Markl, M., Schumacher, M. and Richter, B. (2008). Vocal tract and register changes analysed by real-time MRI in male professional singers-a pilot study. *Logoped. Phoniatr. Vocol.* **33**, 67-73.

Fabre, P. (1957). Un procédé électrique percuntané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence; premiers résultats (A non-invasive electric method for measuring glottal closure during phonation: high frequency glottogr. *Bull. Acad. Nat. Med.* **141**, 66-69.

Fant, G. (1960). *Acoustic Theory of Speech Production*. 's-Gravenhage: Mouton and Co.

Finnegan, E., Luschei, E. and Hoffman, H. (1998). Estimation of alveolar pressure from direct measures of tracheal pressure during speech. *NCVS Status and Progress Report* **12**, 1-10.

Fischer, J., Noser, R. and Hammerschmidt, K. (2013). Bioacoustic Field research: a primer to acoustic analyses and playback experiments with primates. *Am. J. Primatol.* **75**, 643-663.

Fitch, W. T. S. (2002). Primate vocal production and its implications for auditory research. In *Primate Audition - Ethology and Neurobiology* (ed. A. Ghazanfar), pp. 87-108. CRC Press, Inc.

Fitch, W. T. and Hauser, M. D. (1995). Vocal production in nunhuman primates: acoustics, physiology, and functional constraints on "honest" advertisement. *Am. J. Primatol.* **37**, 191-219.

Fitch, W. T., Neubauer, J. and Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Anim. Behav.* **63**, 407-418.

Flanagan, J. (1968). Source-system interaction in the vocal tract. *Ann. N. Y. Acad. Sci.* **155**, 9-17.

Garcia, M., Herbst, C. T., Bowling, D. L., Dunn, J. C. and Fitch, W. T. (2017). Acoustic allometry revisited: morphological determinants of fundamental frequency in primate vocal production. *Sci. Rep.* **7**, 10450.

Gobl, C. and Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Commun.* **40**, 189-212.

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R. and Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *J. Voice* **2**, 118-126.

Green, S. (1975). Variation of Vocal Pattern with Social Situation in the Japanese Moneky (Macaca fuscata): A FieldStudy. In *Primate Behaviour. Developments in Field and Laboratory Research* (ed. L. A. Rosenblum), pp. 1-102. New York: Academic Press.

Green, S. (2010). Dialects in japanese monkeys: vocal learning and cultural transmission of locale-specific vocal behavior? *Z. Tierpsychol.* **38**, 304-314.

Hampala, V., Garcia, M., Švec, J. G., Scherer, R. C. and Herbst, C. T. (2016). Relationship between the electroglottographic signal and vocal fold contact area. *J. Voice* **30**, 161-171.

Henrich, N. (2006). Mirroring the voice from Garcia to the present day: Some insights into singing voice registers. *Log. Phon. Vocol.* **31**, 3-14.

Herbst, C. T. (2016). Biophysics of vocal production in mammals. In *Vertebrate Sound Production and Acoustic Communication* (ed. W. T. Fitch, A. N. Popper and R. A. Suthers), pp. 328. New York: Springer.

Herbst, C. T., Qiu, Q., Schutte, H. K. and Švec, J. G. (2011). Membranous and cartilaginous vocal fold adduction in singing. *J. Acoust. Soc. Am.* **129**, 2253-2262.

Herbst, C. T., Schutte, H. K., Bowling, D. L. and Svec, J. G. (2017). Comparing chalk with cheese - The EGG contact quotient is only a limited surrogate of the closed quotient. *J. Voice* **31**, 401-409.

Herbst, C. T., Lohscheller, J., Švec, J. G., Henrich, N., Weissengruber, G. and Fitch, W. T. (2014). Glottal opening and closing events investigated by electroglottography and super-high-speed video recordings. *J. Exp. Biol.* **217**, 955-963.

Hertegard, S. (2005). What have we learned about laryngeal physiology from high-speed digital videoendoscopy? *Curr. Opin Otolaryngol. Head Neck Surg.* **13**, 152-156.

Herzel, H., Holzfuss, J., Kowalik, Z. J., Pompe, B. and Reuter, R. (1998). Detecting bifurcations in voice signals. In *Nonlinear Analysis of Physiological Data* (ed. H. Kantz, J. Kurths G. Mayer-Kress), pp. 325-344. Berlin: Springer Verlag.

Hirano, M., Kurita, S. and Nakashima, T. (1983). Growth, development, and aging of human vocal folds. In *Vocal Fold Physiology: Contemporary Research and Clinical Issues* (ed. D. Bless), pp. 22-43. San Diego, CA: College Hill Press.

Ishi, C. T., Ishiguro, H. and Hagita, N. (2010). Analysis of the roles and the dynamics of breathy and whispery voice qualities in dialogue speech. *EURASIP Journal on Audio, Speech, and Music Processing* **2010**, 1-12.

Itani, J. (1963). Vocal communication of the wild Japanese monkey. *Primates* **4**, 11-66.

Katsu, N., Nakamichi, M. and Yamada, K. (2016). Function of grunts, girneys and coo calls of Japanese macaques (Macaca fuscata) in relation to call usage, age and dominance relationships. *Behaviour* **153**, 125-142.

Kob, M. (2003) Singing voice modeling - as we know it today. In R. Bresined. Stockhom Music Acoustics Conference, August 6-9, 2003 (SMAC 2003). Stockholm, Sweden, pp. 431-434.

Koda, H. (2004). Flexibility and context-sensitivity during the vocal exchange of coo calls in wild Japanese macaques (Macaca fuscata yakui). *Behaviour* **141**, 1279-1296.

Le Prell, C. G. and Moody, D. B. (1997). Perceptual salience of acoustic features of Japanese monkey coo calls. *J. Comp. Psychol.* **111**, 261-274.

Lohscheller, J. and Eysholdt, U. (2008). Phonovibrogram visualization of entire vocal fold dynamics. *Laryngoscope* **118**, 753-758.

Machida, S. (1990). Threat calls in alliance formation by members of a captive group of Japanese monkeys. *Primates* **31**, 205-211.

Masataka, N. (2010). Motivational referents of contact calls in Japanese monkeys. *Ethology* **80**, 265-273.

Mergell, P., Fitch, W. T. and Herzel, H. (1999). Modelling the role of non-human vocal membranes in phonation. *J. Acoust. Soc. Am.* **105**, 2020-2028.

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H. and Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition* **166**, 84-93.

Orlikoff, R. F. (1991). Assessment of the dynamics of vocal fold contact from the electroglottogram: data from normal male subjects. *J. Speech and Hearing Research* **34**, 1066-1072.

Owren, M. J., Seyfarth, R. M., Dieter, J. A. and Owren, M. J. (1992). “Food” calls produced by adult female rhesus (Macaca Mulatta) and Japanese (M. Fuscata) Macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour* **120**, 218-231.

Pabon, J. P. H. and Plomp, R. (1988). Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *J. Speech Hear. Res.* **31**, 710-722.

Riede, T. (2010). Elasticity and stress relaxation of rhesus monkey (Macaca mulatta) vocal folds. *J. Exp. Biol.* **213**, 2924-2932.

Rothenberg, M. (1981). Acoustic interaction between the glottal source and the vocal tract. In *Vocal Fold Physiology* (ed. K. N. Stevens and M. Hirano), pp. 305-328. Tokyo: University of Tokyo Press.

Rothenberg, M. (1977). Measurement of airflow in speech. *J. Speech Hear. Res* **20**, 155-176.

Rothenberg, M. and Mahshie, J. J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *J. Speech Hear. Res.* **31**, 338-351.

Schneider, B., Zumtobel, M., Prettenhofer, W., Aichstill, B. and Jocher, W. (2010). Normative voice range profiles in vocally trained and untrained children aged between 7 and 10 years. *J. Voice* **24**, 153-160.

Schön Ybarra, M. A. (1995). A comparative approach to the non-human primate vocal tract: implications for sound production. In *Current Topics in Primate Vocal Communication*, pp. 185-198. Boston, MA: Springer US.

Schutte, H. (1980). The Efficiency of Voice Production. (*Doctoral dissertation*), Groningen.

Södersten, M., Hertegård, S. and Hammarberg, B. (1995). Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *J. Voice* **9**, 182-197.

Stathopoulos, E. and Weismer, G. (1985). Oral airflow and air pressure during speech production: a comparative study of children, youths and adults. *Folia Phoniatrica* **37**, 152-159.

Story, B. (2002). An overview of the physiology, physics and modeling of the sound source for vowels. *Acoust. Sci. Tech*. **23**.

Story, B., Hoffman, E. A. and Titze, I. (2003). Vocal tract imaging: a comparison of MRI and EBCT. *Prec. SPIE* **2709**, 209-223.

Sugiura, H. (2008). Vocal exchange of coo calls in Japanese macaques. In *Primate Origins of Human Cognition and Behavior*, pp. 135-154. Tokyo: Springer Japan.

Sulter, A. M., Wit, H. P., Schutte, H. K. and Miller, D. G. (1994). A structured approach to voice range profile (phonetogram) analysis. *J. Speech Hear. Res.* **37**, 1076-1085.

Sundberg, J. (1995). Vocal fold vibration patterns and modes of phonation. *Folia Phoniatr. Logop* **47**, 218-228.

Svec, J. G. and Granqvist, S. (2017). Tutorial and guidelines on measurement of sound pressure level (SPL) in voice and speech. *J. Speech Hear. Res.* (in press) **61**, 441-461.

Taylor, A., Charlton, B. & Reby, D. (2016). Vocal production by terrestrial mammals: source, filter, and function. In *Vertebrate Sound Production and Acoustic Communication* (ed. R. A. Suthers et al.), pp. 229-259. Cham: Springer.

Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *J. Acoust. Soc. Am.* **85**, 901-906.

Titze, I. R. (2000). *Principles of Voice Production*. Iowa City, IA: National Center for Voice and Speech.

Titze, I. R. (2006). *The Myoelastic Aerodynamic Theory of Phonation*. Denver: National Center for Voice and Speech.

Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *J. Acoust. Soc. Am.* **123**, 2733-2749.

Titze, I., Riede, T. and Mau, T. (2016). Predicting Achievable Fundamental Frequency Ranges in Vocalization Across Species. *PLoS Comput. Biol.* **12**, e1004907.

Tokuda, I., Riede, T., Neubauer, J., Owren, M. J. and Herzel, H. (2002). Nonlinear analysis of irregular animal vocalizations. *J. Acoust. Soc. Am.* **111**, 2908-2919.

Van Den Berg, J., (1956). Direct and indirect determination of the mean subglottic pressure. *Folia Phoniatrica* **8**, 1-24.

Van Den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *J. Speech Hear. Res.* **3**, 227-244.

Van Den Berg, J. and Tan, T. S. (1959). Results of experiments with human larynxes. *Pract. Oto-Rhino-Laryng* **21**, 425-450.

Young, R. W. (1939). Terminology for logarithmic frequency units. *J. Acoust. Soc. Am.* **11**, 134-139.