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], Takeshi NISHIMURA[2]

[1] Bioacoustics Laboratory, Department of Cognitive Biology, University Vienna, Althanstrasse 14, 1090 Wien, 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 Dr. Paul Michelon, 42023 Saint-Etienne, France

[*] Corresponding author: herbst@ccrma.stanford.edu

Keywords: Voice production principles, laryngeal configuration, electroglottography, Japanese macaque, voice range profile, excised larynx preparation

Summary statement: *In vivo* and *ex vivo* empirical data of Japanese macaque phonation suggests universal physical and physiological principles of voice production in humans and non-human primates.

Abstract

While the call repertoire and its communicative function is relatively well explored in Japanese macaques (*Macaca fuscata*), little empirical data is 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) was comparable to that of a 7-10 year old human, with the exception of low-intensity chirps, whose production may be facilitated by the species' vocal membranes. In coo calls, redundant control of fundamental frequency in relation to SPL was also comparable to 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.

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 myo-elastic aero-dynamic (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). The latter 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 & Kajiyama 1941; Taylor et al. 2016; Fitch & 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 & Orlikoff 2000), *ex vivo* (Döllinger et al. 2011), and *in silico* (Kob n.d.; Story 2002). In contrast, much less is known about the actual physical and functional/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*). Ever since Itani's groundbreaking work (Itani 1963), the investigation of this species' vocal communication has received wide attention (Le Prell & 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 typically focus 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 will be compared to 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/tracheal air pressure (Schutte 1980; Finnegan et al. 1998), glottal airflow (Rothenberg 1977; Stathopoulos & 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 & Hillman 2010; Lohscheller & 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 Figure 1A). The measured admittance variations are largely proportional to the time-varying 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 is shown in Figure 1B (Berke et al. 1987; Baken & 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 initiated by separation of the lower vocal fold margins;
- e: upper margins start to separate; and
- f: glottis is open, the contact area is at its minimum

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

While 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 & Cannito 1995). Here, we apply EGG data acquisition to *in vivo* phonation of a female Japanese macaque trained to vocalize on command. EGG data is complemented with SPL-calibrated acoustic recordings and matched EGG data from an excised larynx preparation of a

Japanese macaque larynx *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 (#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 ½ year old female Japanese macaque, having a resting vocal fold length of about 7.7 mm, as measured from a CT scan having a spatial resolution of 0.35 x 0.35 mm and a slice interval of 0.2mm.

The animal had been trained over a period of 6 months for another research project (Koda et al. 2017) to sit in a custom-made monkey chair wearing a special purpose jacket (Figure 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 minutes, were conducted over a period of eight days.

EGG signals were recorded with 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 Figure 1A). In this setup, head movement of the animal resulted in intermittent contact loss between the electrodes and the individual's neck in about 60 % of all recorded signals. EGG signals were only considered for further analysis if two conditions were fulfilled: (a) presence of a cyclical EGG signal at a fundamental frequency corresponding to that of the acoustic signal (checked through inspection of respective spectrograms); and (b) 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 ATL SL-8851 sound pressure level meter (ATP Instrumentation Ltd., Leicestershire, UK), applying Method 5 from (Svec & 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 48kHz with a Tascam US-144KMII audio interface (TEAC America Inc., Montebello, CA). 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*

ex vivo data acquisition was conduced at the Department of Cognitive Biology, University of Vienna, Austria. No ethical approval was required. The larynx came from a female Japanese macaque (weight = 7.4 kg, head- body length without tail = 72.6 cm) who 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 elsewhere (Garcia et al. 2017). The resting vocal fold length was visually determined to be about 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 air (100% humidity). 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 electroglottographic recordings (see (Herbst et al. 2014) for details), whilst 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 by the following criteria: (a) comparable fundamental frequency; (b) comparable periodicity and harmonic content (nearly periodic and sinusoidal for coo calls, slightly irregular and slightly aperiodic for growls and chirps; and (c) comparable relative EGG signal level (note that the EGG signal level of “chirp” call was typically about 15 dB to 20 dB lower than that of all other calls – see Figure 3C).

Data analysis

Fundamental frequency (f_o) was estimated with the Praat (Boersma & 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 author CTH: the calibrated sound pressure level (SPL), expressed in dB(C), and the dominant frequency (f_{DOM}) (Fischer et al. 2013), representing the frequency having the maximum amplitude within the analyzed signal portion's acoustic spectrum. The source code of the respective algorithms 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 watts 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 watts per square meter, derived from the measured sound pressure level (SPL @ 30 cm) as

$$I = I_0 10^{SPL/10}, I_0 = 10^{-12} W/m^2 \quad (3)$$

Finally, the aerodynamic power P_{AIR}, expressed in watts, was calculated as the product of the time-averaged glottal air flow and the time-averaged subglottal pressure.

Results

A total of 448 calls were recorded *in vivo*, which were labelled manually according to the classification scheme provided by Green (Green 1975): 377 “coos”, 31 “growls”, 14 “chirps”, and 26 transitions between “coo” and “grunt”. While 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 relation between f_o and SPL for all vocalizations is depicted in Figure 2A. Such a display, called a phonetogram (Damste 1970) or voice range profile (VRP) (Pabon & 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 Figure 2A, allowing for a comparison between the investigated Japanese macaque and humans, are normative VRP data for children aged 7 to 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. (Hirano et al. 1983) – see Figure 2B. A substitution of the vocal fold lengths of the two examined Japanese macaque specimens into the linear regression through Hirano et al's data for children below 12 years of age suggests that comparable vocal fold lengths are approximately found in children aged 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 Figure 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 or greater than 0.8 (39.3 % of all cases) resulted in an average slope of 0.28 semitones / dB SPL. The semitone-scale (Young 1939) was chosen in order for the data to be comparable to 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 about 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 Figure 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 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 Figure 5 for an important counter-example). The EGG signals of the chirps also approximated nearly sinusoidal shapes. However, they had markedly weaker amplitudes (measured as -26.6 dB in Figure 3, as compared to -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 Figure 4. f_o drops abruptly from about 464 Hz to about 190 Hz, while the EGG waveform abruptly alternates between two distinct shapes around t = 280 ms in Figure 4D.

The average harmonics-to-noise ratio (HNR) of all coo calls is plotted against the respective average SPL in Figure 5. The data in panel A suggest an overall trend for HNR to be lower in softer calls. A stereotypical example for 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 acoustic signal of a stereotypical coo call characterized as “non-breathy” (panels D and E) contained 12 harmonics above noise floor, and the corresponding EGG waveform was devoid of visible noise components, resulting in a pronounced waveshape.

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 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 sound pressure levels of targeted call types (recall Figure 2). The supplemented data from the excised larynx experiment allows for the estimation of physical and physiological voice production parameters (recall 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* vs. excised larynx) are linked. While in the current study two different animals' larynges 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, growl, coo, and chirp, had distinct fundamental frequencies and were well separated within the generated phonetogram (Figure 2). The growl and coo calls were well aligned within normative voice range data published for 7 – 10 year old children (Schneider et al. 2010) (but note the greater sound levels of the growl vocalizations in comparison to the respective phonations of children around 200 – 250 Hz). However, even the higher frequencies of the chirps ($f_o \approx 3$ kHz) can be sung by some children of that age, but typically only at high vocal intensities (CTH, personal observation). The VRP comparison is, however, limited by the fact that the VRP data of the children have been acquired via instructions to continuously and fully cover their entire voice range (i.e., reaching the minima and maxima of both sound level and fundamental frequency), 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 indicated by the collected data. Furthermore, while the children's VRP is continuous, the Japanese Macaque's VRP is not, owing to the different methods of data acquisition. It can therefore not 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 Hz 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 fundamental frequency of vocal fold vibration can be approximated with a simple string model as

$$f_o = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}}, \quad (4)$$

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). While 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 assumed to be constant for an individual. Vocal fold length is thus a main anatomical determinant for an individual's fundamental frequency range.

A recent comparative allometric study showed that vocal fold length is a good predictor for the minimum fundamental frequency across eleven 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 about 7.7 mm, and 7.3 mm, respectively. Hirano et al. found comparable vocal fold lengths for children aged about 6 – 10 years (Hirano et al. 1983) - recall Figure 2B. This evidence thus strongly suggests that the similar fundamental frequency ranges of the examined Japanese macaque and 7 to 10 year old children are determined by comparable vocal fold length. This would imply that the string model approximation (Eq. 4) applies to both humans and non-human primates (see also (Riede 2010)), supporting the hypothesis of universal sound production principles.

The similarity between the primate and the human vocal organ is also seen when assessing dynamical aspects of fundamental frequency control. We found an f_o increase of about 0.28 semitones per dB SPL. This value is comparable to data from humans, where an increase of about 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 & Tan 1959; Titze 1989) is a major influence factor for fundamental frequency 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 voice range profile 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 sound levels, a phenomenon which deserves further discussion. We hypothesize that these low sound pressure levels 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 (Tecumseh S. 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 electroglottographic evidence suggested distinct differences in vocal fold vibration patterns for the three call types. The sinusoidal waveforms of the coo calls in Figures 3 and 5C, as well as the chirp call in Figure 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, due to the lack of vocal fold contact.

Interpretation of the other EGG waveforms, including those presented in Figure 3A and Figure 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 Figure 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: (a) the cause for acoustic differences between these call types is clearly laryngeal, similar to different laryngeal mechanisms in human voice registers (Henrich

2006); (b) 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); (c) as expected from a bifurcating system, the two vibratory phenomena do not co-exist.

Some of the softer coo calls had a pronounced "breathy" perceptual quality, as noticed by the experimenters. This potential phenomenon, which is spectrally characterized by fewer noteworthy harmonics and the appearance of high-frequency noise components, was quantified by calculation of the harmonics-to-noise ratio (HNR) – see Figure 5A. The coo calls with lower HNR (see Figure 5B and C for a stereotypical example) typically had only about two to five harmonics above the noise floor. The respective EGG signals assumed a sinusoidal waveshape, with superimposed noise. As suggested above, this is analogous to breathy phonation in 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".

Those "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, as compared to the "breathy" calls. However, as, mentioned above, without clearly established landmarks for EGG signals in Japanese macaques, further interpretation is hazardous.

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/valence states or social communicative context, analogous to what has been shown for humans when using breathy voice in speech (Gobl & Ní Chasaide 2003; Ishi et al. 2010; Miyazawa et al. 2017).

This study has a few limitations that are worth mentioning. For one, 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 & 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.

Conclusion

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 electroglottographic waveforms. The data from this experiment, although coming 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, whose production may be facilitated by the species' vocal membranes. In coo calls, dynamic control of fundamental frequency in relation to sound pressure level was also comparable to humans. Electroglottographic evidence suggested that growls, coos, and chirps were produced by distinct laryngeal vibratory mechanisms, analogous to those of humans. Electroglottographic footage 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 myo-elastic aero-dynamic (MEAD) principle.

Competing interests

No competing interests declared.

Funding

This research has been supported by an ‘‘APART’’ grant received from the Austrian Academy of Sciences (awarded to C. T. Herbst), and by the JSPS-KAKENHI grant no. 16H04848 (awarded to T. Nishimura).

References

Baken, R.J., 1992. Electroglottography. *Journal of Voice*, 6(2), pp.98–110.

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

Beecher, M.D. et al., 2008. Perception of Conspecific Vocalizations by Japanese Macaques. *Brain, Behavior and Evolution*, 16(5–6), pp.443–460. Available at: http://www.karger.com/?doi=10.1159/000121881 [Accessed August 11, 2017].

van den Berg, J., 1956. Direct and indirect determination of the mean subglottic pressure. *Folia Phoniatrica*, 8, pp.1–24.

van den Berg, J., 1958. Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, 3, pp.227–244.

van den Berg, J. & Tan, T.S., 1959. Results of experiments with human larynxes. *Pract.Oto-Rhino-Laryng*, 21, pp.425–450.

Berke, G. et al., 1987. Laryngeal modeling: Theoretical, in vitro, in vivo. *Laryngoscope*, 97, pp.871–881.

Blount, B.G., 1985. "Girney" vocalizations among Japanese macaque females: Context and function. *Primates*, 26(4), pp.424–435. Available at: http://link.springer.com/10.1007/BF02382457 [Accessed August 11, 2017].

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

Bouchet, H., Koda, H. & Lemasson, A., 2017. Age-dependent change in attention paid to vocal exchange rules in Japanese macaques. *Animal Behaviour*, 129, pp.81–92. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0003347217301495 [Accessed September 13, 2017].

Bouhuys, A. et al., 1968. Pressure-flow events during singing. *Annals of the New York Academy of Sciences*, 155, pp.165–176.

Brown, C.H. & Cannito, M.P., 1995. Modes of vocal variation in Syke's monkey (Cercopithecus albogularis) squeals. *Journal of Comparative Psychology*, 109(4), pp.398–415.

Chan, R.W. & Titze, I.R., 2003. Effect of postmortem changes and freezing on the viscoelastic properties of vocal fold tissues. *Ann Biomed Eng*, 31(4), pp.482–491. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12723689.

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

Damste, P.H., 1970. The phonetogram. *Pract Otorhinolaryngol (Basel)*, 32(3), pp.185–187.

Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=5481120.

Deliyski, D.D. & Hillman, R.E., 2010. State of the art laryngeal imaging: research and clinical implications. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 18, pp.147–152.

Döllinger, M. et al., 2011. Experiments on analysing voice production: Excised (human, animal) and in vivo (animal) approaches. *Current Bioinformatics*, 6, pp.286–304.

Echternach, M. et al., 2008. Vocal tract and register changes analysed by real-time MRI in male professional singers-a pilot study. *Logoped Phoniatr Vocol*, 33(2), pp.67–73. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18569645.

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, pp.66–69.

Fant, G., 1960. *Acoustic theory of speech production*, 's-Gravenhage: Mouton and Co.

Finnegan, E., Luschei, E. & Hoffman, H., 1998. Estimation of alveolar pressure from direct measures of tracheal pressure during speech. *National Center for Voice and Speech Status and Progress Report*, 12(June 1998), pp.1–10.

Fischer, J., Noser, R. & Hammerschmidt, K., 2013. Bioacoustic Field Research: A Primer to Acoustic Analyses and Playback Experiments With Primates. *American Journal of Primatology*, 75(7), pp.643–663. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23592340 [Accessed February 24, 2017].

Fitch, W.T. & Hauser, M.D., 1995. Vocal production in nunhuman primates: acoustics, physiology, and functional constraints on "honest" advertisement. *American Journal of Primatology*, 37, pp.191–219.

Fitch, W.T., Neubauer, J. & Herzel, H., 2002. Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, 63, pp.407–418.

Flanagan, J., 1968. Source-system interaction in the vocal tract. *Annals of the New York Academy of Sciences*, 155(1), pp.9–17.

Garcia, M. et al., 2017. Acoustic allometry revisited: morphological determinants of fundamental frequency in primate vocal production. *Scientific Reports*, in review.

Gobl, C. & Ní Chasaide, A., 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, 40(1–2), pp.189–212. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167639302000821 [Accessed August 14, 2017].

Gramming, P. et al., 1988. Relationship between changes in voice pitch and loudness. *Journal of Voice*, 2(2), pp.118–126.

Green, S., 2010. Dialects in Japanese Monkeys: Vocal Learning and Cultural Transmission of

Locale-specific Vocal Behavior? *Zeitschrift für Tierpsychologie*, 38(3), pp.304–314. Available at: http://doi.wiley.com/10.1111/j.1439-0310.1975.tb02006.x [Accessed August 11, 2017].

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

Hampala, V. et al., 2016. Relationship Between the Electroglottographic Signal and Vocal Fold Contact Area. *Journal of Voice*, 30(2), pp.161–171. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0892199715000600 [Accessed March 17, 2017].

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

Herbst, C.T., 2016. Biophysics of Vocal Production in Mammals. In W. T. Fitch, A. N. Popper, & R. A. Suthers, eds. *Vertebrate Sound Production and Acoustic Communication*. New York: Springer, p. 328.

Herbst, C.T. et al., 2017. Comparing chalk with cheese - The EGG contact quotient is only a limited surrogate of the closed quotient. *J Voice*, 31(4), pp.401–409.

Herbst, C.T. et al., 2014. Glottal opening and closing events investigated by electroglottography and super-high-speed video recordings. *J Exp Biol*, 217, pp.955–963.

Herbst, C.T. et al., 2011. Membranous and cartilaginous vocal fold adduction in singing. *J Acoust Soc Am*, 129(4), pp.2253–2262.

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

Herzel, H. et al., 1998. Detecting bifurcations in voice signals. In H. Kantz, J. Kurths, & G. Mayer-Kress, eds. *Nonlinear analysis of physiological data*. Berlin: Springer Verlag, pp. 325–344.

Hirano, M., Kurita, S. & Nakashima, T., 1983. Growth, development, and aging of human vocal folds. In D. Bless, ed. *Vocal Fold Physiology: Contemporary research and clinical issues*. San Diego, CA: College Hill Press, pp. 22–43.

Ishi, C.T., Ishiguro, H. & 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), pp.1–12. Available at: http://asmp.eurasipjournals.com/content/2010/1/528193 [Accessed August 14, 2017].

Itani, J., 1963. Vocal communication of the wild Japanese monkey. *Primates*, 4(2), pp.11–66. Available at: http://link.springer.com/10.1007/BF01659149 [Accessed July 11, 2017].

Katsu, N., Nakamichi, M. & 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(2), pp.125–142. Available at: http://booksandjournals.brillonline.com/content/journals/10.1163/1568539x-00003330 [Accessed August 11, 2017].

Kob, M., Singing voice modeling - as we know it today. In R. Bresin, ed. *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(10), pp.1279–1296. Available at: http://booksandjournals.brillonline.com/content/10.1163/1568539042729685 [Accessed September 13, 2017].

Koda, H., Kunieda, T. & Nishimura, T., 2017. From hand to mouth: Greater effort in motor preparation is required for voluntary control of vocalization than for touching in monkeys. , in review.

Lohscheller, J. & Eysholdt, U., 2008. Phonovibrogram visualization of entire vocal fold dynamics. *Laryngoscope*, 118(4), pp.753–758. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18216742.

Machida, S., 1990. Threat calls in alliance formation by members of a captive group of Japanese monkeys. *Primates*, 31(2), pp.205–211. Available at: http://link.springer.com/10.1007/BF02380942 [Accessed August 11, 2017].

Masataka, N., 2010. Motivational Referents of Contact Calls in Japanese Monkeys. *Ethology*, 80(1–4), pp.265–273. Available at: http://doi.wiley.com/10.1111/j.1439-0310.1989.tb00745.x [Accessed August 11, 2017].

Mergell, P., Fitch, W.T. & Herzel, H., 1999. Modelling the role of non-human vocal membranes in phonation. *Journal of the Acoustical Society of America*, 105(3), pp.2020–2028.

Miyazawa, K. et al., 2017. Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition*, 166, pp.84–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28554088 [Accessed August 14, 2017].

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

Owren, M.J. et al., 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(3), pp.218–231. Available at: http://booksandjournals.brillonline.com/content/journals/10.1163/156853992x00615 [Accessed August 11, 2017].

Pabon, J. & Plomp, R., 1988. Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *Journal of Speech and Hearing Research*, 31, pp.710–722.

Le Prell, C.G. & Moody, D.B., 1997. Perceptual salience of acoustic features of Japanese monkey coo calls. *Journal of Comparative Psychology*, 111(3), pp.261–274.

Riede, T., 2010. Elasticity and stress relaxation of rhesus monkey (Macaca mulatta) vocal folds. *J Exp Biol*, 213(Pt 17), pp.2924–2932. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20709920.

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

Rothenberg, M., 1977. Measurement of airflow in speech. *Journal of Speech & Hearing Research*, 20(1), pp.155–176.

Rothenberg, M. & Mahshie, J.J., 1988. Monitoring vocal fold abduction through vocal fold contact area. *J. Speech and Hearing Research*, 31(September 1988), pp.338–351.

Schneider, B. et al., 2010. Normative Voice Range Profiles in Vocally Trained and Untrained Children Aged Between 7 and 10 Years. *Journal of Voice*, 24(2), pp.153–160. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19303739 [Accessed August 12, 2017].

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*. Boston, MA: Springer US, pp. 185–198. Available at: http://link.springer.com/10.1007/978-1-4757-9930-9_9 [Accessed August 13, 2017].

Schutte, H., 1980. *The efficiency of voice production. (Doctoral dissertation)*, Groningen.

Södersten, M., Hertegard, S. & Hammarberg, B., 1995. Glottal Closure, Transglottal Airflow, and Voice Quality in Healthy Middle-Aged Women. *Journal of Voice*, 9(2), pp.182–197.

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

Story, B., 2002. An overview of the physiology, physics and modeling of the sound source for vowels. *Acoust. Sci. & Tech.*, 23(4).

Story, B., Hoffman, E.A. & Titze, I., 2003. Vocal tract imaging: A comparison of MRI and EBCT. *Preceeding from SPIE*.

Sugiura, H., 2008. Vocal Exchange of Coo Calls in Japanese Macaques. In *Primate Origins of Human Cognition and Behavior*. Tokyo: Springer Japan, pp. 135–154. Available at: http://www.springerlink.com/index/10.1007/978-4-431-09423-4_7 [Accessed August 11, 2017].

Sulter, A.M. et al., 1994. A structured approach to voice range profile (phonetogram) analysis. *Journal of Speech and Hearing Research*, 37, pp.1076–1085.

Sundberg, J., 1995. Vocal fold vibration patterns and modes of phonation. *Folia Phoniatrica et Logopaedica*, 47, pp.218–228.

Svec, J.G. & Granqvist, S., 2017. Tutorial and guidelines on measurement of sound pressure level (SPL) in voice and speech. *Journal of Speech and Hearing Research*, in press.

Taylor, A., Charlton, B. & Reby, D., 2016. Vocal Production by Terrestrial Mammals: Source, Filter, and Function. In R. A. Suthers et al., eds. *Vertebrate Sound Production and Acoustic Communication*. Cham: Springer, pp. 229–259.

Tecumseh S. Fitch, W., 2002. Primate Vocal Production and Its Implications for Auditory Research.

In A. Ghazanfar, ed. *Primate Audition - Ethology and Neurobiology*. CRC Press, Inc., pp. 87–108. Available at: http://www.crcnetbase.com/doi/abs/10.1201/9781420041224.ch6 [Accessed August 13, 2017].

Titze, I., Riede, T. & Mau, T., 2016. Predicting Achievable Fundamental Frequency Ranges in Vocalization Across Species F. E. Theunissen, ed. *PLOS Computational Biology*, 12(6), p.e1004907. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27309543 [Accessed February 6, 2017].

Titze, I.R., 2008. Nonlinear source-filter coupling in phonation: theory. *J Acoust Soc Am*, 123(5), pp.2733–2749. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18529191.

Titze, I.R., 1989. On the relation between subglottal pressure and fundamental frequency in phonation. *J Acoust Soc Am*, 85(2), pp.901–906. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2926005.

Titze, I.R., 2000. *Principles of Voice Production*, National Center for Voice and Speech.

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

Tokuda, I. et al., 2002. Nonlinear analysis of irregular animal vocalizations. *J Acoust Soc Am*, 111(6), pp.2908–2919. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12083224.

Young, R.W., 1939. Terminology for Logarithmic Frequency Units. *Journal of the Acoustical Society of America*, 11, pp.134–139.

Figures

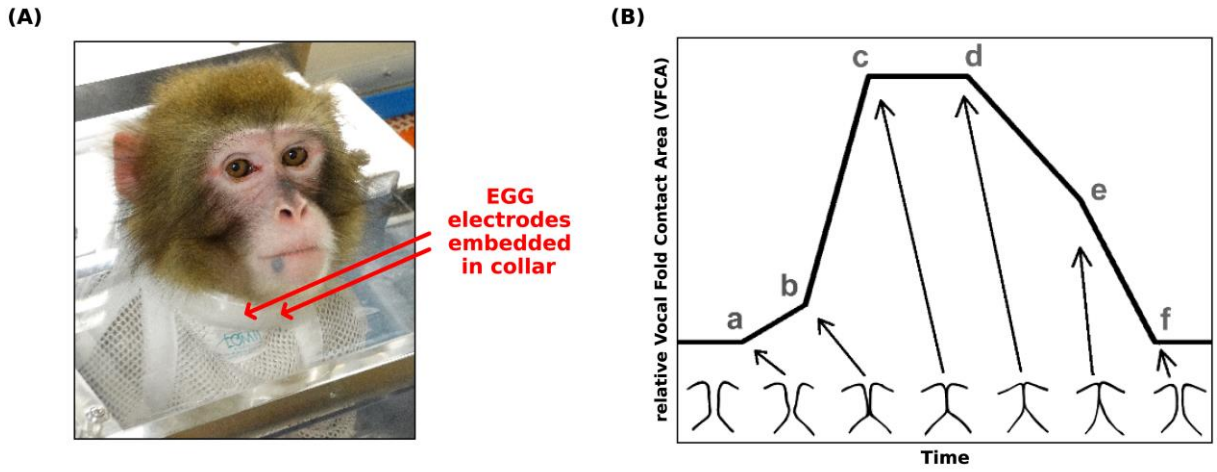

Figure 1: ***Electroglottography (EGG)****. (A) Attachment of EGG electrodes in experimental setup; (B) Schematic illustration of EGG waveform for one glottal cycle (Baken & Orlikoff 2000; Berke et al. 1987), with illustrations of vocal fold movement and contact within the coronal plane shown at the bottom (see text).*

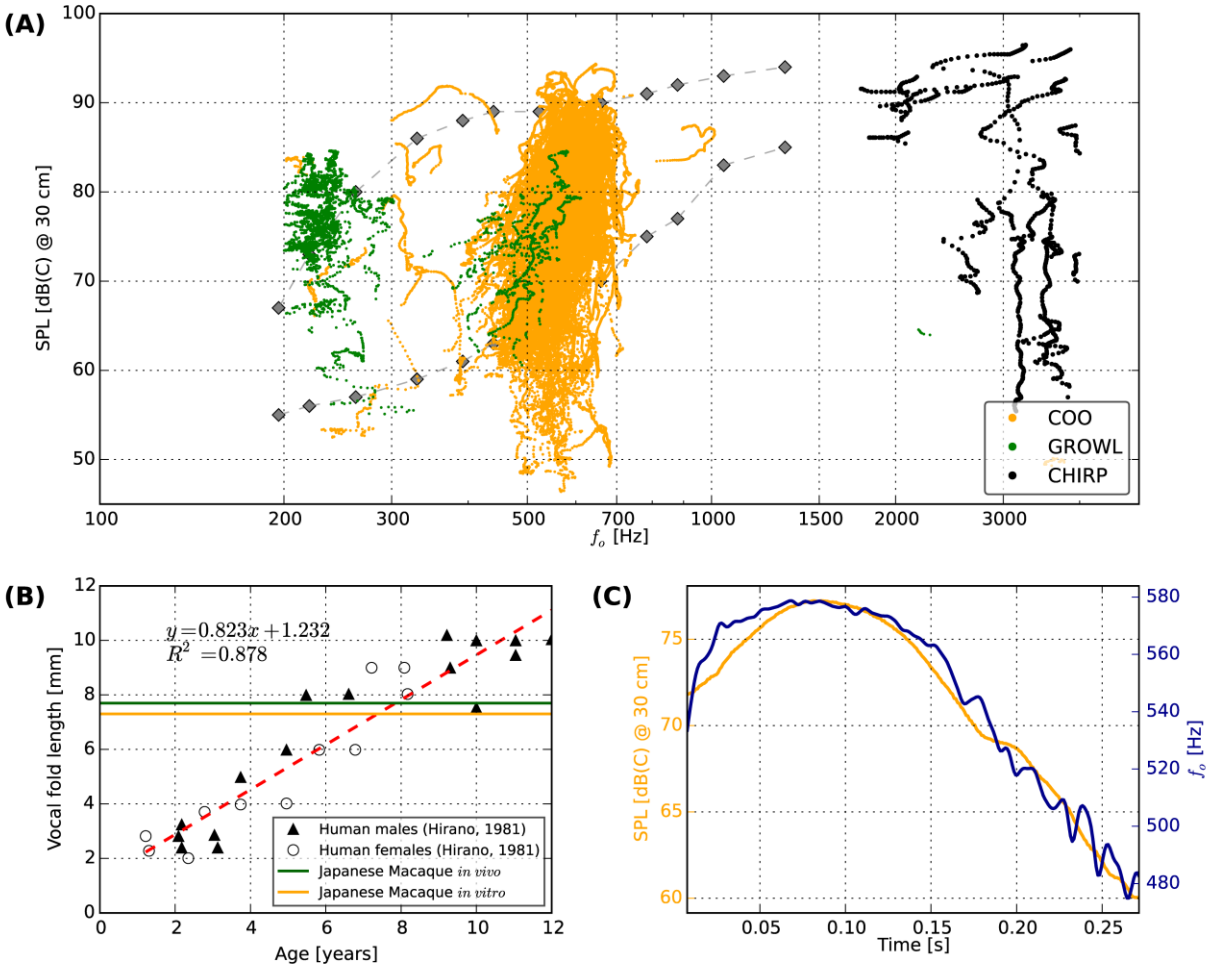

Figure 2: ***Fundamental frequency (f_o) and sound pressure level (SPL) of analyzed vocalizations****. (A) Phonetogram, showing f_o vs SPL. The superimposed diamonds and dashed lines represent normative voice range data from human children aged 7 to 10 years (Schneider et al. 2010); (B) Vocal fold length measures for pre-pubertal children according to Hirano et al. (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.*

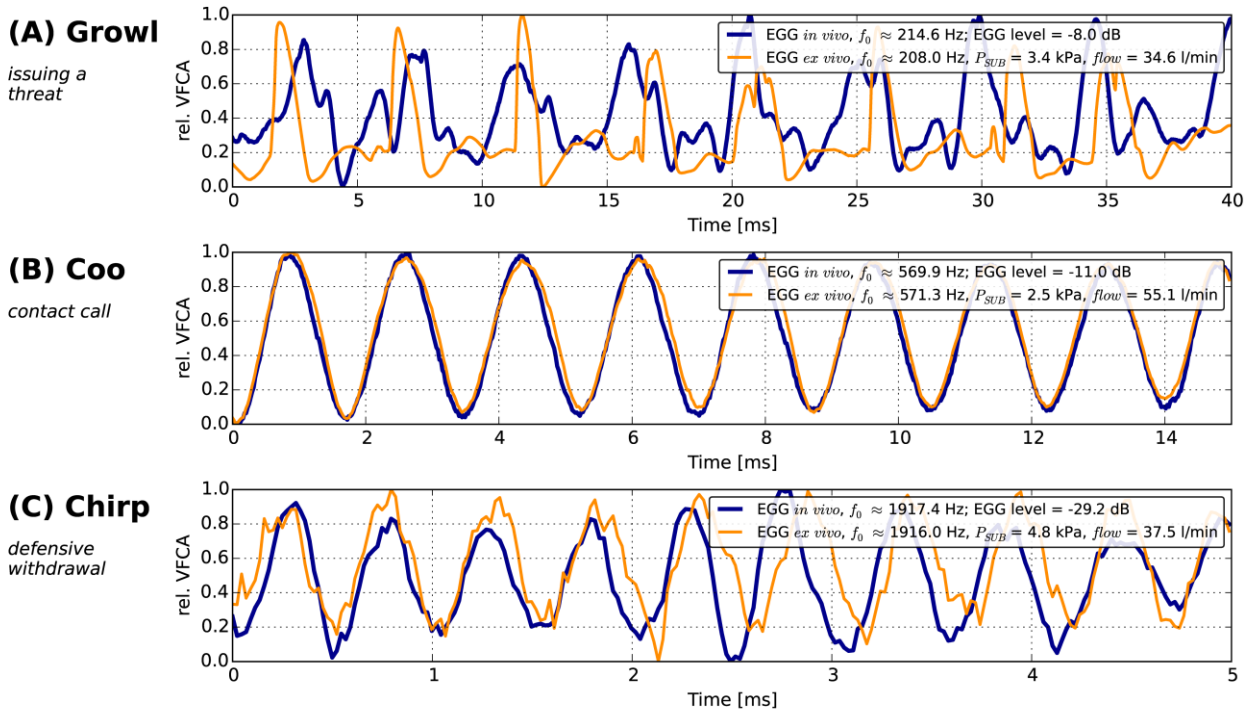

Figure 3: ***Comparable EGG waveforms of* in vivo *and excised larynx recordings for all three call types.***

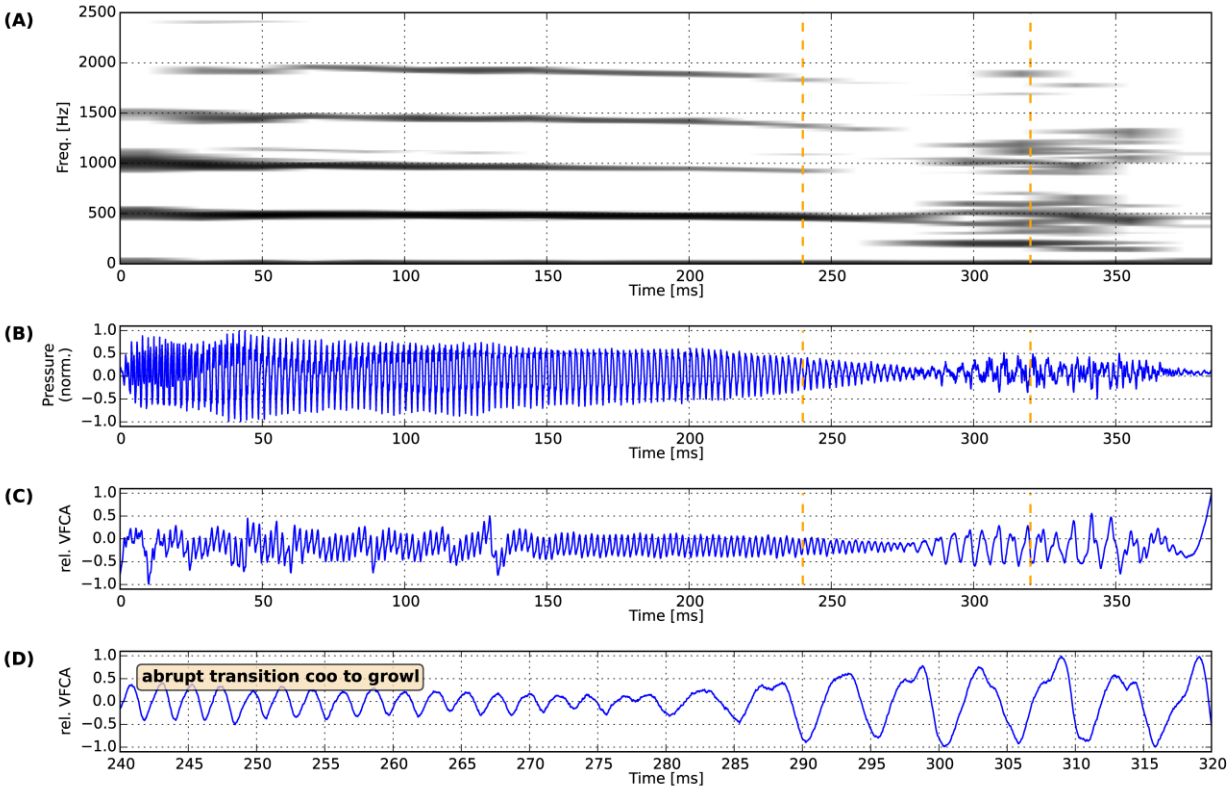

Figure 4: ***Abrupt transition from "coo" to lower frequency "growl"****. (A) Narrow-band spectrogram of microphone signal; (B) acoustic signal; (C) EGG signal; (D) portion of the EGG signal, extracted at t = 240 – 320 ms.*

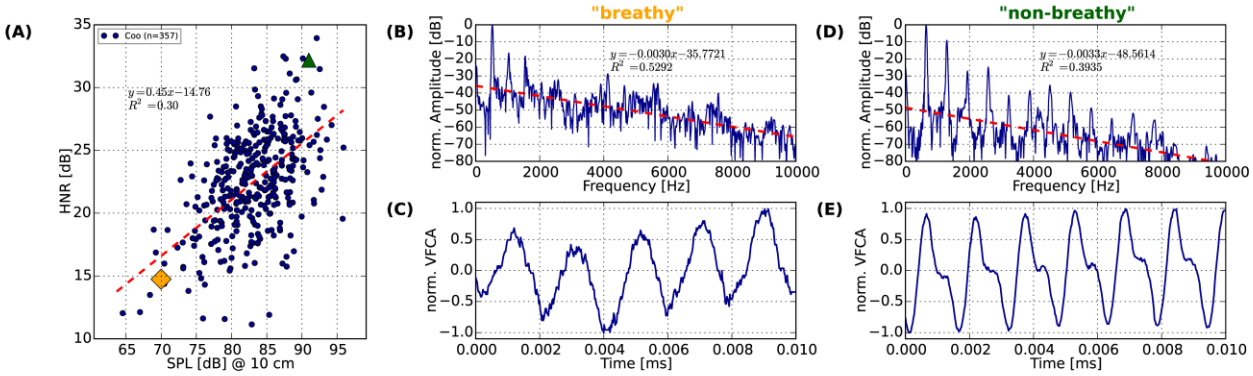

Figure 5: ***Variation in spectral quality between "coo" calls****. (A) Average harmonics-to-noise ratio (HNR) as a function of average SPL per call (one blue circle represents one coo call). The calls identified with an orange diamond and a green triangle are described in more detail in panels (B) –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

Call Type	Number of Calls	Number of Data points	Mean fo [Hz]	Mean fDOM [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	127981	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)

Table 1: *Mean* f_o, f_{DOM}, SPL, and HNR data and standard deviations *for all call types.*

Call type	PSU B [kPa]	air flow [l/min]	SPL [dB(C) @ 30 cm]	EGL [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

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