The descended larynx and the descending larynx

Takeshi NISHIMURA1*

1Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan

Received 20 November 2017; accepted 1 March 2018

Abstract Our understanding of the evolution of human speech has been expanded by an increased knowledge of vocal anatomy and physiology in non-human primates. Comparative approaches provide evidence supporting the primate origins of many speech faculties. The descent of the larynx enables the two-tube configuration of the supralaryngeal vocal tract (SVT) in humans; however, this configuration is also found in chimpanzees and macaques. The acoustic properties of voices produced in helium gas support the view that vocalizations are usually produced through SVT resonance, with the sound source generated by vibration of the vocal folds in gibbons and marmosets, as seen in human speech. Non-human primates produce a wider range of vocal repertoire than previously thought, reflecting their varied manipulations of the vocal apparatus to modify SVT topology. These species often actively descend the hyoid and larynx to produce calls. This ‘active’ descent is one of the options for SVT modification in non-human primates. However, this is distinct from human speech, where a ‘static’ descended larynx moves in a restricted range during speech. Instead, humans modify SVT configuration by combinations of contraction and relaxation of the tongue muscles, to produce their vocal acoustics. The components of the vocal apparatus act under the constraint of anatomy, and various associations of anatomy and vocal actions are expected to be found in a variety of types of vocalization in non-human primates. Increasing knowledge of their anatomy and physiology promises better understanding of primate origins and of the evolutionary history of physical faculties in human speech.

Key words: source–filter theory, supralaryngeal vocal tract, tongue, descent of the larynx, speech evolution

Introduction

The origin of language remains one of the most enigmatic issues for understanding human evolution. Paleoanthropologists have continued to challenge this issue, which undoubtedly contributed to the unfolding of humanity and its civilizations (Nishimura, 2008). Unfortunately, their efforts have faced a major obstacle: languages do not fossilize.

No human population lacks verbal communication in the form of speech. Humans have physical faculties enabling them to produce many distinct phonemes sequentially, including vowels and consonants, even in a short single exhalation (Greenberg et al., 2003). Speech is a kind of vocalization and it is not per se the same as language in a narrow sense. Nevertheless, speech has made major evolutionary contributions to the origin and evolution of the languages with which we are now endowed, by providing a medium that is efficient and suitable for language communication (Fitch et al., 2005). The prominent acoustic properties of speech are achieved through sophisticated manipulation of the vocal apparatus, including the lungs and pulmonary apparatus as the power source; the larynx for phonation to generate the sound source; and the tongue, jaw, and lips for articulation to generate the radiated acoustic phenomenon constituting voice (Lieberman, 1984; Titze, 1994; Fitch, 2000a).

The source–filter theory explains well the acoustic and physiological mechanisms of articulation in speech production (Chiba and Kajiyama, 1941; Fant, 1960; Titze, 1994). The radiated sound is induced by the airflow exhaled in response to the compression of pulmonary volume. It is regulated by the voluntary contraction or relaxation of thoracic muscles, including the diaphragm, and by passive recoil forces within the pulmonary apparatus. The exhaled airflow runs up through the glottis, to induce cyclical vibrations of the vocal folds (VF) to generate the sound source. The vibration is induced passively by airflow from the lung and results in self-sustained oscillation, described by myoelastic aerodynamics (van den Berg, 1958; Titze, 1980). The acoustic properties of the source are characterized by the degree of periodicity and by the harmonic structure of the generated sound, i.e. the fundamental frequency ($f_0$) and its higher harmonics. The relationships between the sub- and supraglottal pressures, in addition to the physical properties of VF, determine source quality, which affects the intensity, volume, and pitch of voices. The physical processes are complicated and some of them remain under debate (Herbst, 2016). The sound source makes the air filling the SVT resonate in
accordance with its resonance property, which is dictated by the volumetric topology of the SVT. The resonance property is also influenced by the speed of sound, which is influenced by the air density, temperature, and pressure. This means that the SVT serves as a filter to amplify some harmonics of the source near to its inherent resonance frequencies, and to suppress others. The voiced sounds with some bands of the amplified harmonics—formants—are radiated from the mouth. The distribution pattern of the formant positions determines the kind of voiced sound, such as vowels, that we perceive (Fant, 1960; Stevens, 1972; Titze, 1994).

Here, current knowledge of the vocal anatomy and physiology of articulation in non-human primates is surveyed to argue for varied vocal physiology, within the restrictions of their given anatomy, as the cradle of speech evolution.

Vocal anatomy in non-human primates

Some primates show anatomical modifications to existing apparatus for producing species-distinct vocalizations. Howler monkeys, species in the genus *Alouatta* in Central and South America, are well known to have an enlarged and modified hyolaryngeal complex, which is believed to contribute to producing their noisy and low-frequency roars (Starck and Schneider, 1960; Schön, 1971; Dunn et al., 2015). The laryngeal air sac is also often cited for its contributions to acoustics (Fitch, 2000a). Five forms of the laryngeal sac are found in non-human primates, though this feature is lacking in certain species such as gibbons, and is also absent in humans (Starck and Schneider, 1960; Hayama, 1970; Hewitt et al., 2002). Siamangs (*Symphalangus syndactylus*) from Southeast Asia have a large laryngeal air sac that opens to the laryngeal cavity just above the glottis (Mott, 1924; Némai and Kelemen, 1933; Starck and Schneider, 1960; Hayama, 1970). This sac is inflated just before they start their loud calls (Marshall and Marshall, 1976). Sac inflation is also found in other species, such as putty-nosed guenons (*Cercopithecus nictitans*) (Itani, 1963). These inflatable sacs are probably used to amplify the voice sound rather than to modify its formant pattern (Riede et al., 2008; de Boer, 2009). This view is supported by the finding that the removal of the sac reduces loudness in putty-nosed guenons (Gautier, 1971; Gautier and Gautier, 1977). While this may be true in some species, future examinations might uncover novel acoustic contributions of this feature. Nevertheless, the sac hinders the production of the distinct voices required for speech (de Boer, 2009).

The vocal membranes or vocal lips are also suggested to make acoustic contributions in many species of non-human primates (Negus, 1949). These features are thin upward extensions from the VFs and probably lower the subglottal pressure, resulting in increased efficiency in producing loud and high-pitched calls (Mergell et al., 1999; Fitch, 2000a).

Varied repertoires of vocalizations are partly dependent on evolutionary modifications to the anatomy of existing vocal apparatus in non-human primates.

Descent of the larynx in humans and non-human primates

Humans also have specific anatomical features in terms of the SVT and tongue (Lieberman, 1984; Fitch, 2000a). The SVT is composed of horizontal oral and vertical pharyngeal cavities. The two cavities are almost equally long and are located almost perpendicular to each other, and the channel between the two cavities—the oropharyngeal isthmus—is narrowed in humans. By contrast, the oral is longer than the pharyngeal cavities without a narrowed isthmus in non-human primates, and the tongue is also globular in humans but flat in non-human primates (Laitman and Reidenberg, 1997; Nishimura, 2005; Nishimura et al., 2008; Takemoto, 2008). This two-tube configuration of the human SVT with a globular tongue provides anatomical efficiency for sequential and varied modifications of the SVT topology (Lieberman, 1968; Lieberman, 1984; Takemoto, 2001). This feature develops in infant and early juvenile periods. Human babies show an SVT configuration and tongue posture rather similar to those seen in non-human primates. The static positions of the hyoid and larynx are close to the soft palate and the epiglottis is located in the nasopharynx, at an ‘intranarial’ position (Negus, 1949; Sasaki et al., 1977). The hyoid and larynx descend along the neck, and the epiglottis is detached from the soft palate in the first 9 years of human life (Sasaki et al., 1977; Fitch and Giedd, 1999; Lieberman et al., 2001). This descent of the larynx lengthens the pharyngeal cavity faster than changes occur in the oral cavity, and pulls the tongue base into the pharynx to develop the two-tube SVT and globular tongue in human adults.

Non-human primates also demonstrate descent in the static position of the larynx during growth. Chimpanzees (*Pan troglodytes*) show laryngeal descent with detachment of the epiglottis from the soft palate, as seen in humans (Nishimura et al., 2003, 2006; Nishimura, 2005). Their pharynx faces the posterior surface of the tongue, but is shorter than in humans. Such developmental descent of the larynx is also found in Japanese macaques, while their epiglottis remains at an intranarial position even in adults (Laitman et al., 1977; Flügel and Rohen, 1991; Nishimura et al., 2008). In fact, their laryngeal skeleton is almost articulated to the hyoid, while both are connected by a moderately long ligament and certainly act independently from each other in hominoids including humans (Nishimura, 2003). Thus, the pattern of descent of the larynx varies among non-human primates.

Regardless of this varied descent of the larynx, humans and non-human primates have distinctly different configurations of the SVT. The two-tube configuration is prevented from forming in non-human primates by facial growth as well as by the short descent of the larynx. Their faces continue to grow long after the larynx has descended, and they develop a long oral cavity and a flat tongue (Nishimura, 2005; Nishimura et al., 2006, 2008). By contrast, growth of the human face is almost complete in the early juvenile stage and remains short and flat even in adults (Fitch and Giedd, 1999; Lieberman et al., 2001). This keeps the oral cavity short and makes the tongue globular.

The descent of the larynx is associated with developmental changes in swallowing mechanisms during early infancy.
and with the development of speech at early juvenile stages in humans (Sasaki et al., 1977; Lieberman, 1984; Fitch, 2000a). In human babies, liquids and food pass through the bilateral channels of the laryngeal opening—the piriform recesses—while the epiglottis is held at an intranarial position (Negus, 1949; Sasaki et al., 1977). This configuration changes to the adult mode in late infancy, where food and liquid boluses pass over the ventral surface of the epiglottis, which bends back over the laryngeal opening (Sasaki et al., 1977; Ekberg and Sigurjónnsson, 1982). This adult mode of swallowing is also found in adult macaques (Larson and Herring, 1996) and is probably a consequence of developmental changes as seen in humans (Crompton et al., 1997). Whereas swallowing physiology and its developmental patterns are yet to be examined in the other non-human primates, varied descent of the larynx among non-human primates is considered to interact with developments in other physiological functions rather than vocalizations, such as swallowing as seen in human infants.

Vocal physiology in non-human primates

Varied vocal repertoires are found in non-human primates (McCombs and Semple, 2005). The particular type of acoustic traits in some of them might be expected to depend on a distinct vocal physiology. The physiological mechanisms of animal vocalizations are sometimes examined by studying the acoustics of voices in helium gas, which are modified from natural acoustics (Nowicki, 1987; Rand and Dudley, 1993; Ballintijn and ten Cate, 1998; Yamada and Okanoya, 2003; Madsen et al., 2012; Reber et al., 2015; Pasch et al., 2017). Under helium-enriched conditions, the speed of sound is increased and the resonance frequencies of the SVT are also shifted upward (Nowicki, 1987). For vocalizers using SVT filtering, the acoustic features of their voices are inevitably influenced by such increased sound speed and all formants dependent on the SVT resonances are shifted upward in frequency (Nowicki, 1987; Rand and Dudley, 1993).

Helium experiments also distinguish the nature of the source–filter interaction. Source–filter independence is another important feature of speech physiology. This means that the property of the sound source is only slightly influenced by the resonance of the SVT (Chiba and Kajiyama, 1941; Fant, 1960; Titze, 1994). The $f_0$ position is changeable independently from the SVT acoustics in human speech. This is in contrast to the rigid source–filter interactions. If such interactions were to exist, it would imply that vibrations of the VF’s are inevitably influenced by the acoustics of the SVT, where SVT resonances primarily determine the $f_0$ position (Fletcher and Rossing, 1998). Such a strong interaction is seen in some extreme cases of human singing and in some wind instruments, e.g. trombone. It would prevent us from producing a series of distinct voiced sounds with a steady pitch in speech. Thus, source–filter independency is one of the requirements for human speech in terms of flexible modifications of vocal tone.

In the case of source–filter independence, the helium-enriched atmosphere does not shift $f_0$, whereas only formants are shifted. By contrast, in the case of a strong source–filter interaction, it should shift $f_0$ upward to a similar degree as the formants. Studies of helium-modified voices provide evidence suggesting that: (i) there are few filtering effects in the vocal sac that is inflated under the jaw in frogs (Rand and Dudley, 1993); (ii) the source–filter independence applies to song birds (Nowicki, 1987); (iii) and moderate source–filter interactions apply to grasshopper mice (Onychomys) in North America (Pasch et al., 2017).

Studies of non-human primates to date all support the idea that their vocalizations are explained by the source–filter theory and independence, as seen in human speech. White-handed gibbons (Hylobates lar) from Southeast Asia produce loud, high-pitched, puretone-like, and melodious calls, termed ‘songs.’ This puretone-like voice is produced in accordance with source–filter independence, different from wind instruments (Koda et al., 2012). This source is thought to be produced by tuning the first formant with $f_0$, as in human soprano singing (Koda et al., 2012), and future empirical evidence is expected to support such an idea. Common marmosets (Callithrix jacchus) from South America produce high-pitched ‘phee’ calls with an $f_0$ of >7000 Hz (Bezerra and Souto, 2008). These whistle-like voices are also produced in accordance with source–filter independency (Koda et al., 2015). These physiological manipulations can be performed by humans, but they are only rarely performed during speech. Whereas vocal physiology has so far been revealed in only these two species among non-human primates, these studies strongly suggest that varied vocal repertoires are produced by differences of manipulation in vocal physiology and anatomy which are universal to primates, including humans, rather than by diversifications of physiological mechanisms on demand (Koda et al., 2012, 2015).

Anatomy and physiology in non-human primates

Non-human primates probably perform varied manipulations in vocal physiology. Gibbons regulate the source property and the filter effects of the SVT principally independently from each other (Koda et al., 2012). Their melodious songs are achieved by highly coordinated modifications in the source property of the VF vibration and filtered by resonance in the SVT (Koda et al., 2012). While the direct observation is expected, this suggests that increasing the vibration rate to shift $f_0$ upward is synchronized with changes in the SVT topology (probably by shortening its length and/or enlarging the mouth opening) to shift upward the first resonance frequency of the SVT, to maintain a formant tuning with pitch. Such sophisticated manipulation is probably achieved by neural control of the actions of the peripheral components of the vocal apparatus, rather than by mechanical associations between them. Diana monkeys (Cercopithecus diana) from West Africa demonstrate a formant transition in that the formant pattern changes in a single alarm call (Riede and Zuberbühler, 2003; Riede et al., 2005). Putty-nosed guenons from West Africa combine distinct calls into a third structure having another meaningful message, which is different from each of the original calls (Arnold and Zuberbühler, 2006; Zuberbühler, 2006). Campbell’s monkeys (Cercopithecus campbelli cambelli) from West Africa combine an alarm call followed by a suffix call to broaden the meaning of the original call (Ouattara et al.,...
2009). While such combinations are often argued in a term of protosyntax, the formant transitions and combinations are achieved by sequential modifications of the SVT topology, strongly suggesting coordinated actions of the components of the vocal apparatus (Riede et al., 2005). Macaques from Asia and restricted parts of North Africa are species that have been well surveyed in vocal repertoire, and probably have the ability to modify their SVT topology and produce a range of formant patterns larger than thought previously (Fitch et al., 2016). Thus, recent empirical studies have challenged the traditional view that non-human primates generate a narrow repertoire of formants. Such an underestimated stems from overemphasizing the distinctions in vocal anatomy between humans and non-human primates.

Advanced faculties in human speech are not similar to vocalizations in any other terrestrial mammals. The globular tongue is one of the anatomical foundations for dynamic modifications of the two-tube SVT topology (Lieberman, 1968; Takemoto, 2001), and the flat tongue physically prevents non-human primates from performing the tongue actions shown by humans (Lieberman, 1968; Takemoto, 2008). This anatomical distinction has been overemphasized, and the low plasticity of SVT modifications is believed to be primarily responsible in non-human primates (Lieberman et al., 1969). However, recent empirical studies have demonstrated the composite and dynamic acoustics of their vocalizations (Riede and Zuberbühler, 2003; Arnold and Zuberbühler, 2006; Zuberbühler, 2006; Ouattara et al., 2009), strongly indicating that these animals have the physical faculties to implement dynamic modifications of their SVT topology better than thought previously (Riede et al., 2005; Koda et al., 2012; Fitch et al., 2016).

Non-human primates and other mammals are often found to lower the larynx actively during vocalizations (Fitch, 2000b; Fitch and Reby, 2001). This ‘active’ descent of the larynx lowers the positions of the resonance frequencies by extending the SVT length from the glottis to the mouth (Fitch and Reby, 2001). Such active descent also occurs during human singing, to produce ‘singing formants’ that emphasize resonance in the posterior pharyngeal cavity (Sundberg, 1974). This action is one of the options to modify SVT topology, by extending the pharyngeal cavity and pulling the tongue base downward into the pharynx to modify tongue posture (Fitch et al., 2016). It should be noted that the ‘static’ descended larynx in humans moves in a rather restricted range during speech production (Hiemae et al., 2002; Hiemae and Palmer, 2003). Such small actions are not attributed to any anatomical restrictions, and the hyoid and larynx move in a larger range for mastication, and upward for swallowing (Ekberg and Sugiourjonsson, 1982; Hiemae et al., 2002; Hiemae and Palmer, 2003). In speech, the hyoid–laryngeal complex moves slightly anteriorly and remains there as an anchor for the tongue. The globular tongue changes its surface topology per se by combinations of contraction and relaxation in each of the tongue muscles in the manner of muscular hydrostats (Smith and Kier, 1989; Takemoto, 2001). While the SVT topology is probably modified by various dynamic actions of the vocal apparatus, these actions in non-human primates are not always the same as those performed by humans during speech.

Conclusions

Evolutionary modifications in anatomy have exerted major permanent influences on physiological functions involving a given apparatus. The components of the vocal apparatus are involved in basic and vital physiological functions other than vocalizations, including the lungs and pulmonary apparatus for respiration, the larynx for preventing involuntary aspiration, and the tongue, jaws, and lips for mastication and swallowing. Anatomical modifications in these regions have been limited by potential risks to their functions. By contrast, temporary actions of the existing vocal apparatus exert a restricted and temporal influence on the other functions, suggesting that they can be adopted readily with advantages for vocalization. Nevertheless, such functions are inevitably defined by the anatomy of a given apparatus. Non-human primates probably perform various dynamic actions in their vocal apparatus for generating their wide vocal repertoire, within the limits of their given anatomy. Therefore, advancing the technology for examining the anatomy and physiology of vocalizations in non-human primates is expected to enhance our understanding of evolutionary history and of the constraints underlying the evolution of speech in human lineages.

Acknowledgments

This work was supported by SPIRITS program from Kyoto University and by a Grant-in-Aid for Scientific Research (#16H04848). I gratefully thank the editor and anonymous reviewers for comments on an earlier version of the manuscript.

References

Arnold K. and Zuberbühler K. (2006) Language evolution: semantic combinations in primate calls. Nature, 441: 303.
Ballintijn M.R. and ten Cate C. (1998) Sound production in the collared dove: a test of the ‘whistle’ hypothesis. Journal of Experimental Biology, 201: 1637–1649.
Bezerra B.M. and Souto A. (2008) Structure and usage of the vocal repertoire of Calithrix jacchus. International Journal of Primatology, 29: 671–701.
Chiba T. and Kajiya M. (1941) The Vowel: Its Nature and Structure. Tokyo-Kaiseikan, Tokyo.
Crompton A.W., German R.Z., and Thexton A.J. (1997) Mechanisms of swallowing and airway protection in infant mammals (Sus domesticus and Macaca fascicularis). Journal of Zoology, 241: 89–102.
de Boer B. (2009) Acoustic analysis of primate air sacs and their effect on vocalization. Journal of the Acoustical Society of America, 126: 3329–3343.
Dunn J.C., Halenar L.B., Davies T.G., Cristobal-Azkarate J., Reby D., Sykes D., Dengg S., Fitch W.T., and Knapp L.A. (2015) Evolutionary trade-off between vocal tract and testes dimensions in howler monkeys. Current Biology, 25: 2839–2844.
Ekberg O. and Sugiourjonsson S.V. (1982) Movement of the epiglottis during deglutition. A cineradiographic study. Gastroenteral Radiology, 7: 101–107.
Fant G. (1960) Acoustic Theory of Speech Production. Mouton, The Hague.
Fitch W.T. (2000a) The evolution of speech: a comparative review. Trends in Cognitive Sciences, 4: 258–267.
Fitch W.T. (2000b) The phonetic potential of nonhuman vocal
tructs: comparative cineradiographic observations of vocalizing animals. Phonetica, 57: 205–218.
Fitch W.T. and Giedd J. (1999) Morphology and development of the human vocal tract: a study using magnetic resonance imaging. Journal of the Acoustical Society of America, 106: 1511–1522.
Fitch W.T. and Reby D. (2001) The descended larynx is not uniquely human. Proceedings of the Royal Society of London: B. Biological Science, 268: 1669–1675.
Fitch W.T., Hauser M.D., and Chomsky N. (2005) The evolution of the language faculty: clarifications and implications. Cognition, 97: 179–210.
Fitch W.T., de Boer B., Mathur N., and Ghazanfar A.A. (2006) Monkey vocal tracts are speech-ready. Science Advances, 2: e1600723.
Fletcher N.H. and Rossing T.D. (1998) The Physics of Musical Instruments. Springer, New York.
Flügel C. and Rohen J.W. (1991) The craniocaudal proportions and laryngeal position in monkeys and man of different ages (a morphometric study based on CT-scans and radiographs). Mechanisms of Ageing and Development, 61: 65–83.
Gautier J.P. (1971) Etude morphologique et fonctionnelle des anexes extra-laryngées des cercopithécines; liaison avec les cris d’espacement. Biologia Gabonica, 7: 229–267.
Gautier J.P. and Gautier A. (1977) Communication in old world monkeys. In: Seboek T.A. (ed.), How Animals Communicate. Indiana University Press, Bloomington, pp. 890–964.
Greenberg S., Carvey H., Hitchcock L., and Chang S.Y. (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. Journal of Phonetics, 31: 465–485.
Hayama S. (1970) The saccus laryngis in primates. Journal of the Anthropological Society of Nippon, 78: 274–298 (in Japanese with English abstract).
Herbst C.T. (2016) Biophysics of vocal production in mammals. In: Suthers R.A., Fitch W.T., Fay R.R., and Popper A.N. (eds.), Vertebrate Sound Production and Acoustic Communication. Springer Handbook of Auditory Research, Vol. 53. Springer International Publishing, Cham, pp. 159–189.
Hewitt G., MacLarnon A., and Jones K.E. (2002) The functions of laryngeal air sacs in primates: a new hypothesis. Folia Primatologica, 73: 70–94.
Hiemae K.M. and Palmer J.B. (2003) Tongue movements in feeding and speech. Critical Reviews in Oral Biology and Medicine, 14: 413–429.
Hiemae K.M., Palmer J.B., Medicis S.W., Hegener J., Jackson B.S., and Lieberman D.E. (2002) Hyoid and tongue surface movements in speaking and eating. Archives of Oral Biology, 47: 11–27.
Itani J. (1983) Vocal communication of the wild Japanese monkey. Primates, 4: 11–66.
Koda H., Nishimura T., Tokuda I.T., Oyakawa C., Nihonmatsu T., and Masataka N. (2012) Soprano singing in gibbons. American Journal of Physical Anthropology, 149: 347–355.
Koda H., Tokuda I.T., Wakita M., Ito T., and Nishimura T. (2015) The source-filter theory of whistle-like calls in mammals: acoustic analysis and simulation of helium-modulated voices. Journal of the Acoustical Society of America, 137: 3068–3076.
Laitman J.T. and Reidenberg, J.S. (1997) The human aerodigestive tract and gastrointestinal reflux: an evolutionary perspective. American Journal of Medicine, 103: 2S–8S.
Laitman J.T., Crelin E.S., and Conlogue G.J. (1977) The function of the epiglottis in monkey and man. Yale Journal of Biology and Medicine, 50: 43–48.
Larson J.E. and Herring S.W. (1996) Movement of the epiglottis in mammals. American Journal of Physical Anthropology, 100: 71–82.
Lieberman P. (1968) Primate vocalizations and human linguistic ability. Journal of the Acoustical Society of America, 44: 1574–1584.
Lieberman P. (1984) The Biology and Evolution of Language. Harvard University Press, Cambridge.
Lieberman P.H., Klatt D.H., and Wilson W.H. (1969) Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. Science, 164: 1185–1187.
Lieberman D.E., McCarthy R.C., Hiemae K.M., and Palmer J.B. (2001) Ontogeny of postnatal hyoid and larynx descent in humans. Archives of Oral Biology, 46: 117–128.
Madsen P.T., Jensen F.H., Carder D., and Ridgway S. (2012) Dolphin whistles: a functional misnomer revealed by heliox breathing. Biology Letters, 8: 211–213.
Marshall J.T., Jr and Marshall E.R. (1976) Gibbons and their terri- torial songs. Science, 193: 235–237.
McCorm K. and Semple S. (2005) Coevolution of vocal communication and sociality in primates. Biology Letters, 1: 381–385.
Mergell P., Fitch W.T., and Herzel H. (1999) Modeling the role of nonhuman vocal membranes in phonation. Journal of the Acoustical Society of America, 105: 2020–2028.
Mott F. (1924) A Study by serial sections of the structure of the larynx of Hylobates syndactylus (Siamang gibbon). Proceedings of the Zoological Society of London, 94: 1161–1170.
Negus V.E. (1949) The Comparative Anatomy and Physiology of the Larynx. William Heinemann Medical Books, London.
Némai J. and Kelemen G. (1933) Beiträge zur Kenntnis des Gibbonkehlkopfes. Zeitschrift für Anatomie und Entwicklungs geschichte, 100: 512–520.
Nishimura T. (2003) Comparative morphology of the hyo-laryngeal complex in anthropoids: two steps in the evolution of the descent of the larynx. Primates, 44: 41–49.
Nishimura T. (2005) Developmental changes in the shape of the supralaryngeal vocal tract in chimpanzees. American Journal of Physical Anthropology, 126: 193–204.
Nishimura T. (2008) Understanding the dynamics of primate vocalization and its implications for the evolution of human speech. In: Masataka N. (ed.), The Origins of Language. Springer, Tokyo, pp. 111–131.
Nishimura T., Mikami A., Suzuki J., and Matsuzawa T. (2003) Descent of the larynx in chimpanzee infants. Proceedings of the National Academy of Sciences of the United States of America, 100: 6930–6933.
Nishimura T., Mikami A., Suzuki J., and Matsuzawa T. (2006) Descent of the hyoid in chimpanzees: evolution of face flattening and speech. Journal of Human Evolution, 51: 244–254.
Nishimura T., Oishi T., Suzuki J., Matsuda K., and Takahashi T. (2008) Development of the supralaryngeal vocal tract in Japanese macaques: implications for the evolution of the descent of the larynx. American Journal of Physical Anthropology, 135: 182–194.
Novicki S. (1987) Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. Nature, 325: 53–55.
Ouattara K., Lemasson A., and Zuberbühler K. (2009) Campbell’s monkeys use affiliation to alter call meaning. PLoS One, 4: e7808.
Pasch B., Tokuda I.T., and Riede T. (2017) Grasshopper mice employ distinct vocal production mechanisms in different social contexts. Proceedings of the Royal Society of London. Series B, Biological Sciences, 284: 1158.
Rand A.S. and Dudley R. (1993) Frogs in helium—the anuran vocal sac is not a cavity resonator. Physiological Zoology, 66: 793–806.
Reber S.A., Nishimura T., Janisch J., Robertson M., and Fitch W.T. (2015) A Chinese alligator in heliox: formant frequencies in a crocodilian. Journal of Experimental Biology, 218: 2442–2448.
Riede T. and Zuberbühler K. (2003) The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. Journal of the Acoustical Society of America, 114: 1132–1142.
Riede T., Bronson E., Hatzikirou H., and Zuberbühler K. (2005)
Vocal production mechanisms in a non-human primate: morphological data and a model. Journal of Human Evolution, 48: 85–96.

Riede T., Tokuda I.T., Munger J.B., and Thomson S.L. (2008) Mammalian laryngeal air sacs add variability to the vocal tract impedance: physical and computational modeling. Journal of the Acoustical Society of America, 124: 634–647.

Sasaki C.T., Levine P.A., Laitman J.T., and Crelin E.S., Jr (1977) Postnatal descent of the epiglottis in man: a preliminary report. Archives of Otolaryngology, 103: 169–171.

Schön M.A. (1971) The anatomy of the resonating mechanism in howling monkeys. Folia Primatologica, 15: 117–132.

Smith K.K. and Kier W.M. (1989) Trunks, tongues, and tentacles: moving with skeletons of muscle. American Scientist, 77: 29–35.

Starck D. and Schneider R. (1960) Respirationsorgane, A. Larynx. In: Hofer H.O., Schultz A.H., and Starck D. (eds.), Primatologia, Vol. 3, Part 2. Karger, Basel, pp. 423–587.

Stevens K.N. (1972) The quantal nature of speech: evidence from articulatory-acoustic data. In: David E.E., Jr and Denes P.B. (eds.), Human Communication: A United View. McGraw-Hill, New York, pp. 51–66.

Sundberg J. (1974) Articulatory interpretation of the ‘singing formant.’ Journal of the Acoustical Society of America, 55: 838–844.

Takemoto H. (2001) Morphological analyses of the human tongue musculature for three-dimensional modeling. Journal of Speech, Language, and Hearing Research, 44: 95–107.

Takemoto H. (2008) Morphological analyses and 3D modeling of the tongue musculature of the chimpanzee (Pan troglodytes). American Journal of Primatology, 70: 966–975.

Titze I.R. (1980) Comments on the myoelastic-aerodynamic theory of phonation. Journal of Speech and Hearing Research, 23: 495–510.

Titze I.R. (1994) Principles of Voice Production. Prentice Hall, Englewood Cliffs.

van den Berg J. (1958) Myoelastic-aerodynamic theory of voice production. Journal of Speech and Hearing Research, 1: 227–244.

Yamada H. and Okanoya K. (2003) Song syntax changes in Bengalese finches singing in a helium atmosphere. Neuroreport, 14: 1725–1729.

Zuberbühler K. (2006) Language evolution: the origin of meaning in primates. Current Biology, 16: R123–125.