

Vocal Tract Modeling Techniques: From Human Voice to Non-Human Primates Vocalizations

M. Gamba¹, V. Torti¹, C. Colombo¹, C. Giacomini¹

¹ Department of Life Sciences and Systems Biology, Via Accademia Albertina 13, 10123 Torino, Italy.
E-mail: marco.gamba@unito.it

KEY WORDS: acoustics, phonation, evolution, resonance, formants, comparative approach.

Introduction

The Source-Filter Theory (Fant, 1960) offered a powerful frame for the interpretation of human vocal production. Two main events take place in the human apparatus during vocal production. The first happens at a glottal level. Vocal fold vibration generates the fundamental frequency of a voice and its harmonics. These characteristics are known collectively as the voice source. The source is then modified by the shape and length of the vocal tract. The vocal tract acts as a resonator, adjusting the relative intensities of the frequencies of the source. The column of air vibrates in a complex manner that is influenced by the length and the shape of the vocal tract. One or several resonant frequencies of the vocal tract correspond to prominent spectral peaks called "formants". The position and variation of the formants have been found to have a significant impact on the way humans recognize speech sounds. Even if a model of vocal production based on the relationship between the vocal tract area function and the formant output has been the most common framework for understanding speech production in humans, the study of vocal tract resonance in non-human primates has not comparably developed. This is probably due to several reasons. First, the study of formants in non-human primates started as an attempt to demonstrate that primates could not produce human speech sound. Once the pioneering studies of Lieberman and colleagues (Lieberman, 1968; 1969; Lieberman *et al.*, 1969; 1972) showed that the anatomy and morphology of the non-human primate vocal tract prevented the production of human-like sounds, this field of investigation immediately ceased its activity. Some years later, the work of Andrew (1976) and then Hauser (Hauser *et al.*, 1993; Hauser, 1996) brought back some attentions to the meaning of formants in primate intra-specific communication. In more recent years, a number of studies have shown that formant-based semantic communication is also present in non-human primates, for instance in Diana monkey alarm calls (Riede and Zuberbühler, 2003 a, b; Rendall *et al.*, 2005). These findings were strengthened when it was found that macaques could, without training, discriminate differences in the formant

structure of their conspecific calls (Fitch and Fritz, 2006). Thus, it is now widely accepted that the calls of many non-human primates, and mammals in general (Taylor and Reby, 2010), possess formants.

A further extension of the importance of vocal tract filtering in primates is the application of computational models to describe their phonation processes. From an acoustic and physiological point of view, human vocal communication is far better known than any other mammal communication system, and techniques from speech science have often been applied to the study of vocal production in other mammals, especially non-human primates (Riede *et al.*, 2005; Gamba and Giacomini, 2006). The purpose of this paper is to introduce a framework for future studies of the relation between vocal-tract shape and acoustics in human and non-human primates. Showing the potential of using vocal tract modeling in non-human primates, we highlight differences and similarities compared to vocal tract modeling in humans.

Materials and Methods

Various imaging techniques are now available for researchers to derive information about the configuration and the morphology of the vocal tract. The two most common options are magnetic resonance imaging (MRI) and x-ray computed tomography (CT). They have the advantage of high resolution in image acquisition and they are potentially useful to study non-human primates too. However they are rarely available in non-human primates range countries, and they may not contrast properly organs of partially de-frozen specimens (Gamba, unpublished data). Because of this, data used in this paper come from different sources. On one side, we are using vocal tract area functions of human vocal tracts available in the VTAR program default vocal tract configurations (Zhou *et al.*, 2001). On the other side, we will use vocal tract area functions determined by measuring cross-sectional areas with 1 cm increment from the vocal folds toward mouth opening and nostrils. These measures were taken over a silicon cast of the entire vocal tract (glottis to lips for the oral tract and glottis to nostrils for the nasal tract respectively) of a large *Lemur catta* male, whose cadaver belonged to the collection of dead animals at the Parc Botanique et Zoologique Tsimbazaza, Antananarivo,

Madagascar. All length and dimension measurements of the cast were taken with a Mitutoyo digital caliper (accurate to 0.01 mm). We measured an average diameter because the cross-section of the vocal tract cast was not generally circular and cross-sectional areas were computed starting from these measures in Microsoft Excel. In both human and non-human primate vocal tract models, computational representations of the vocal tract were built using concatenated tube models [a system comprising a series of concatenated tubes of fixed length, each showing a specific cross-sectional area] (Story *et al.*, 1996; Gamba and Giacoma, 2006). Cross-sectional areas were used to build the cross-sectional area function that represents the input of the vocal tract modeling software VTAR, a Matlab-based computer program for vocal tract acoustic response calculation (Zhang and Espy-Wilson, 2004). VTAR is capable of simulating complex frequency-domain models for the vocal-tract acoustic response where the vocal tract is decomposed into various modules such as simple tubes, branching, and lateral channels. We did two separate models for the oral tract and the nasal tract on the basis of the assumption that non-human primates produce

vocalizations through the nose, or the mouth, but not both (Fitch, 2000). For human nasal resonance models, we used a couple area of 1.04 cm (default value) and we modeled the tract using one nostril to shorten calculation times.

Results

Computational models of the vocal tract simulate the morphology of a specific configuration of the vocal apparatus and attempt to calculate the resonance properties of its cavities. The models used real data about the vocal tract morphology to virtually recreate it, and hence calculate the resonance frequencies that would be observed from the original emitter. We constructed models representing the vibrating air cavity in the vocal tract between the glottis and the lips for two typical sounds of human speech /AA/ and /M/, and those of *Lemur catta* in two species-typical utterances, 'tonal call' and 'grunt'. Fig. 1a shows the vocal tract area function of a typical male /AA/ with a vocal tract length of 17 cm. Fig.

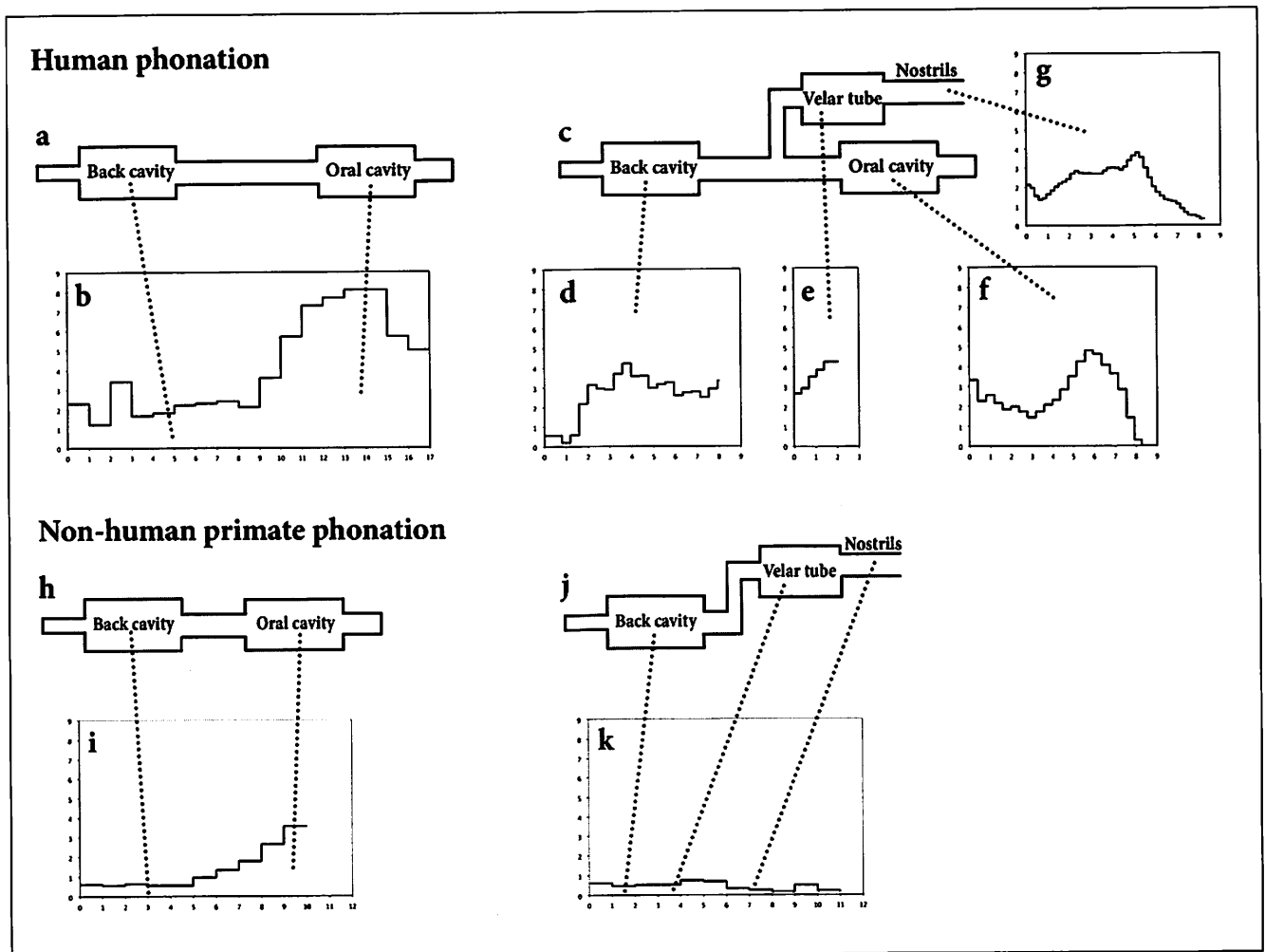


Fig. 1. Human phonation: (a) schematic representation of the human vocal tract during phonation through the oral tract; (b) vocal tract area function during the emission of /AA/; (c) schematic representation of the human vocal tract during phonation through the nasal and the oral tracts; vocal tract area functions during the emission of /M/: (d) back cavity, (e) velar tube, (f) oral cavity, (g) nostrils. Non-human primate phonation: (h) schematic representation of the vocal tract during phonation through the oral tract; (i) vocal tract area function during the emission of a tonal call in *L. catta*; (j) schematic representation of the vocal tract during phonation through the nasal tract; (k) vocal tract area functions during the emission of a nasal grunt in *L. catta*. Schemes are not in scale. Diagrams show distance from the origin (e.g. glottis) in cm on the x-axis and vocal tract area in cm² on the y-axis.

1b shows the vocal tract area function of a male ring-tailed lemur 'tonal call' with a vocal tract length of 10 cm. Both human's /AA/ and lemur's 'tonal call' are emitted with open mouth, with the column of air resonating in the oral tract. The main differences between humans and non-human primates lie in the interaction between the nasal and the vocal tract. The human speech apparatus is a complex system and the full model of the vocal tract for a typical nasal sound like /M/ should include various branches: the back cavity, the velar tube, the oral cavity and one or two nostrils. Figure 1 shows the vocal tract area function of a typical male /M/ with a length of 7.99 cm for the back cavity (d), 2.04 cm for the velar tube (e), 8.66 cm for the oral tract (f) and 8.25 cm for the nostrils (g). In the typical /M/ of the human voice the column of air resonates both in the oral and in the nasal cavities. Fig. 1h shows the vocal tract area function of a male ring-tailed lemur nasal 'grunt' with a vocal tract length of 11 cm. Lemur's 'tonal call' are emitted with open mouth with the column of air resonating in the oral tract.

Acoustic response of the vocal tract models is shown in Fig. 2. The longer, L-shaped tube of a typical human male produces lower resonance frequencies in both simulations (Fig. 2a and c). For the ring-tailed lemur, the slightly longer nasal tract and the narrow, spongy chambers of the nasal airways produce a remarkably lower first formant (F1, Fig. 2d) when compared with that showed by the simulation of the oral tract resonance (Fig. 2b). Third (F3), fourth (F4) and fifth (F5) formants also showed lower frequencies; meanwhile second formant (F2) has an invariant location across the two simulations (Fig. 2b and 2d).

Discussion

In despite of the difference between humans and non-human primates, which anatomically lies in the morphology of the upper airways and in the interaction between the nasal and the vocal tracts, the ability of using formants to extract information related to the physical characteristics (e.g. species, sex, body size, identity) of the emitter may be one functional link between the vocalizations of human and nonhuman primates. Vocal tract models attempt to simulate the behaviour of sound waves in the vocal apparatus and are particularly useful to understand the resonance properties of a specific vocal tract configuration. Understanding the morphology of the vocal tract framework is a crucial step in mapping resonance properties of both humans and animals. It is also a necessary step toward quantitative modeling of species-specific abilities, which requires accurate dimensions of all structures of the vocal apparatus.

Output from the simulations produced by VTAR can be used to gain an understanding of the differences in the application of computational modeling of the vocal tract in humans and in non-human primates, and vocal tract modeling in general may represent a powerful tool to investigate differences and similarities in the encoding of communication sounds in humans and in non-human animals.

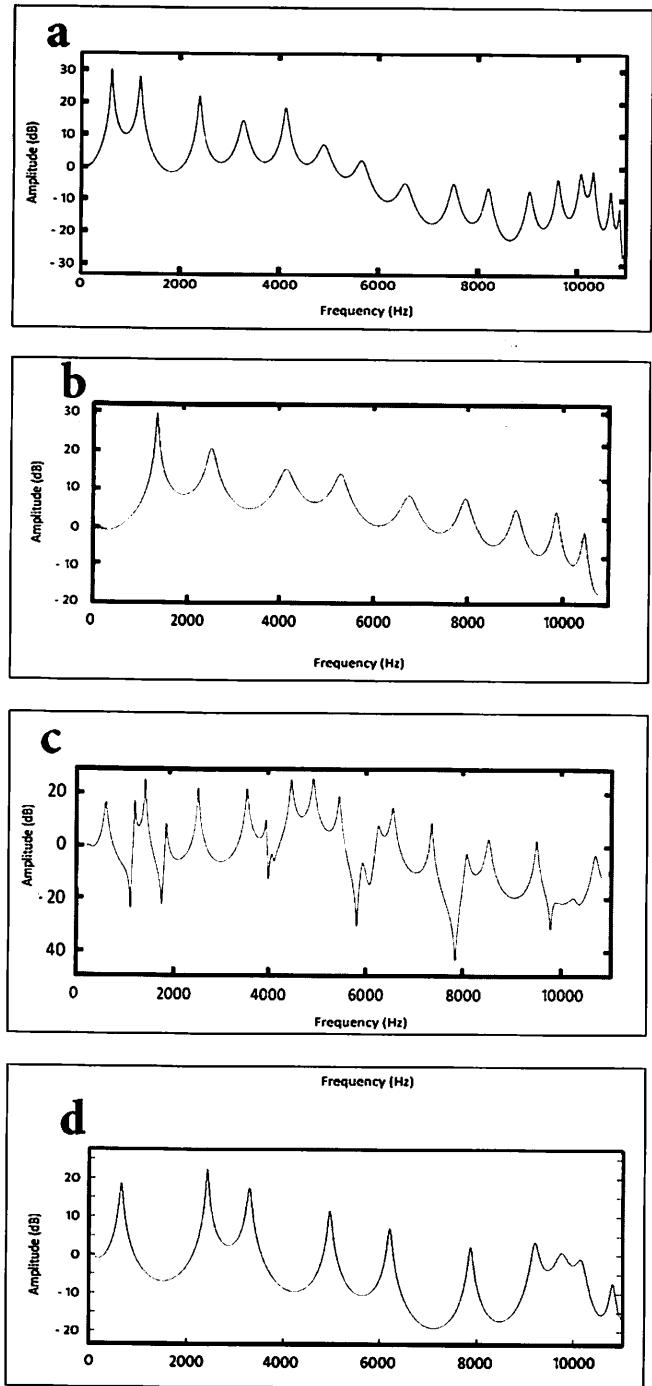


Fig. 2. Acoustic response of the vocal tract models. First 5 formants calculated for the emission of (a) /AA/ in a standard male human vocal tract: 671 Hz, 1245 Hz, 2440 Hz, 3327 Hz, 4182 Hz; (b) a tonal call in *L. catta*: 1223 Hz, 2422 Hz, 4050 Hz, 5245 Hz, 6750 Hz; (c) /MI/ in a standard male human vocal tract: 410 Hz, 1029 Hz, 1697 Hz, 2380 Hz; (d) a nasal grunt in *L. catta*: 652 Hz, 2431 Hz, 3303 Hz, 4964 Hz, 6206 Hz. Diagrams show resonance frequency in Hz on the x-axis and amplitude on the y-axis.

References

- Andrew R.J. 1976. Use of Formants in the Grunts of Baboons and other Nonhuman Primates. *Ann. (N.Y.) Acad. Sci.*, 280: 673-679.
 Fant G. 1960. *Acoustic Theory of Speech Production*. Mouton, the Hague.

- Fitch W.T. 2000. The phonetic potential of nonhuman vocal tracts: comparative cineradiographic observations of vocalizing animals. *Phonetica*, 57: 205-218.
- Fitch W.T., Fritz J.B. 2006. Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J. Acoustic. Soc. Am.*, 120: 2132-2141.
- Hauser M.D., Evans C.S., Marler P. 1993. The Role of Articulation in the Production of Rhesus Monkey (*Macaca mulatta*) Vocalizations. *Anim. Behav.*, 45: 423-433.
- Hauser M., 1996. *The evolution of communication*. MIT Press, Cambridge, MA.
- Gamba M., Giacoma C. 2006. Vocal tract modeling in a prosimian primate: the black and white ruffed lemur. *Acta Acust. United Ac.*, 92: 749-755.
- Lieberman P. 1968. Primate vocalizations and human linguistic ability, 44:1574-1784. (Reprinted in *Perspectives on Human Evolution*. 2. Washburn and Dolhinow (eds.) New York: Holt, Reinhart and Winston; and in *The Speech Code*, Mattingly and Engelhardt (eds.), Sinauer).
- Lieberman P. 1969. On the acoustic analysis of primate vocalizations. *Behavior Research Methods and Instruments*, 1: 169-174.
- Lieberman P.H., Klatt D.H., Wilson W.H. 1969. Vocal Tract Limitations on the Vowel Repertoires of Rhesus Monkey and other Nonhuman Primates. *Science*, 164: 1185-1187.
- Lieberman P., Harris K.S., Wolff P., Russell L.H. 1972. Newborn infant cry and nonhuman-primate vocalizations. *J. Speech Hearing Res.*, 14: 718-727.
- Rendall D., Kollias S., Ney C., Lloyd P. 2005. Pitch (F0) and Formant Profiles of Human Vowels and Vowel-Like Baboon Grunts: the Role of Vocalizer Body Size and Voice-Acoustic Allometry. *J. Acoustic. Soc. Am.*, 117 (2): 944-955.
- Riede T., Zuberbühler K., 2003a. Pulse Register Phonation in Diana Monkey Alarm Calls. *J. Acoustic. Soc. Am.*, 113: 2919-2926.
- Riede T., Zuberbühler K. 2003b. The Relationship Between Acoustic Structure and Semantic Information in Diana Monkey Alarm Vocalization. *J. Acoustic. Soc. Am.*, 114: 1132-1142.
- Riede T., Bronson E., Hatzikirou H., Zuberbühler K. 2005. Vocal production mechanisms in a non-human primate: morphological data and a model. *J. Human Evol.*, 48: 85-96.
- Story B.H., Titze I.R., Wong D. 1996. A simplified model for simulation and transformation of speech. *Proc. of the IEEE Joint Symposia on Systems and Intelligence*, Rockville, MD. 4-5 Nov.
- Taylor A.M., Reby D. 2010. Contribution of the source-filter theory to the study of mammal vocal communication. *J. Zool.*, 280: 221-236.
- Zhang Z., Espy-Wilson C. 2004. A Vocal Tract Model of American English /L/. *J. Acoustic. Soc. Am.*, 115: 1274-1280.
- Zhou X., Zhang Z., Espy-Wilson C. 2001. VTAR: A Matlab-based computer program for vocal tract acoustic modelling. *J. Acoustic. Soc. Am.*, 115: 2543-2543.