Vocal tract proportions and the evolution of speech: New data to answer old questions

Steven Moran[1,2], Sam Kirkham[3], Daniel Friedrichs[2,4],
David S. Strait[5,6,7], Axel G. Ekström[8]

[1] *Institute of Biology, University of Neuchâtel, Switzerland*
[2] *Department of Anthropology, University of Miami, US*
[3] *Department of Linguistics & English Language, Lancaster University, UK*
[4] *Department of Computational Linguistics, University of Zürich, Switzerland*
[5] *Department of Anthropology, Washington University in St. Louis, US*
[6] *Palaeo-Research Institute, University of Johannesburg, South Africa*
[7] *DFG Center for Advanced Studies "Words, Bones, Genes, Tools,"*
University of Tübingen, Germany
[8] *Speech, Music & Hearing, KTH Royal Institute of Technology, Sweden*
steven.moran@unine.ch, s.kirkham@lancaster.ac.uk, daniel.friedrichs@uzh.ch,
dstrait@wustl.edu, axeleks@kth.se

Abstract

In evolutionary phonetics research, important distinctions between humans and human ancestors are contingent on observations that modern humans possess "roughly 1:1" proportions between horizontal and vertical vocal tract sections. However, few descriptions of diverse adult speaker samples exist. We measured horizontal distance between the anterior-most point of the lower lip and the posterior wall of the pharynx, and vertical distance between staphylion and the superior margin of the posterior lamina of the cricoid cartilage. Our results show that proportions exhibit moderate variation between speakers. We discuss these findings in the context of the purported speech capacities of human ancestors.

Introduction

Modern humans are born with "monkey-like" vocal tracts. That is, children are born with flat tongues contained almost wholly in the oral cavity, high larynges and narrow pharynges (Crelin, 1973; Bosma, 1975; Kent, 1981; Kent & Vorperian, 1995) – a set of vocal tract features that closely resemble those of extant non-human primates (Negus, 1949; Lieberman et al., 1969, 1972, 1992; Nishimura, 2005; Takemoto, 2008). In comparison, the development of the adult human vocal tract is characterized by a suite of morphological changes from infancy to adulthood, including descent of the larynx and hyoid, and partial descent of the tongue root into the pharynx, with the vocal tract bending at its midpoint to form a right angle in the nasopharyngeal region. The "adult" vocal tract is attained only throughout the course of maturation, around the ages of 6-8 (Lieberman & McCarthy, 1999; Fitch & Giedd, 1999; Lieberman et al., 2001; Vorperian et al., 2005, 2009).

Literature on the evolution of speech capacities in the human lineage, beginning as early as Victor Negus' (1949) writings on the comparative anatomy of the airways. Negus hypothesized that the uniquely human vocal tract configuration reflected an evolutionary selection pressure for improved speech communication. Yet, this anatomy contributes to an elevated risk of choking (as food may get lodged in the trachea). Thus, the positive benefits of the human vocal tract configuration must have offset the negative effect of the increased risk of choking, implying that the human condition represents an evolutionary

tradeoff (Negus, 1949; Lieberman, 2012, 2017; Ekström, 2024). In this view, the human vocal tract became differentiated from that of its closest extant relatives "… for purposes of speech" (Negus, 1949, p. 198). Would-be benefits to speech production include attaining "optimal" configurations for the full range of speech sounds (Carré et al., 1995, 2017; de Boer, 2010a, 2010b; but see Badin et al., 2014), and in the view of Lieberman (Lieberman et al., 1972; Lieberman, 2012) facilitating the production of "quantal" vowels (Stevens, 1972, 1989) and the robustness of consonant distinctions (Liberman et al., 1967; Lieberman et al., 1992). Vocal tract reconfiguration would also have resulted in an increased ability to readily close off the velum, facilitating non-nasalized phonemic contrasts (e.g., Lieberman, 1991).

In light of this line of argument, it is surprising that while extensive investigations have sought to uncover mechanisms of maturation *per se*, and the developmental trajectories of a variety of vocal tract structures (Crelin, 1973; Bosma, 1975; Lieberman & McCarthy, 1999; Fitch & Giedd, 1999; Lieberman et al., 2001; Vorperian et al., 2005, 2009), there are few descriptions of the variation of inter-structure relationships in diverse samples of adult speakers. Specifically, providing descriptions of variations of such measurements is necessary to either support, refute, or nuance central tenets of the dominant model of the evolution of speech anatomy (Lieberman, 2012; Ekström, 2024). Here, we provide such a description.

Methods

Vocal tract data

We measured horizontal and vertical supralaryngeal vocal tract lengths (SVT_H, SVT_V) and proportions from the University from the Southern California Speech and Vocal Tract Morphology MRI Database (Sorensen et al., 2017; Lim et al.,

2021). The dataset contains real-time magnetic resonance images (rtMRI) of vocal tract shapes observed during read and spontaneous speech from a diverse speaker sample (N=75, 40 females). The speaker sample was aged 18-59 years of age. The sample consists of 49 native and 26 non-native American English speakers (29 females). We used data from neutral/resting positions, in order to avoid any distortion from real-time speech.

Figure 1. SVT_H and SVT_V were measured linearly.

Measurement procedure

SVT_H was measured as the horizontal distance between the anteriormost point on the lower lip and the posterior wall of the pharynx (composed primarily of the constrictor muscles). SVT_V was measured as vertical distance between staphylion (the posteriormost point of the hard palate, typically discernable as the convergence of the void related to cortical bone, the opaque cancellous bone within the palate, and the superior border of the soft palate) and the superior margin of the posterior lamina of the cricoid cartilage (often, although not always, discernable as a slender vertically oriented oval anterior to the vertebral column; the vocal ligament arises from the vocal process of the arytenoid cartilages that sit above the posterior lamina, so this point approximates the superiormost portion of the vocal ligament, see Figure 1).

Results

Of the 75 available subjects, 55 (27 females) were acceptable for our study. Images from the rest were either deemed too poor in quality or had the subject in (qualitatively) highly flexed or highly extended neck postures. Our measurements are likely affected by subjects being in a supine position while the rtMRI data was collected. Our results showed a moderate degree of inter-subject variation with regard to SVT_H-SVT_V proportions. We also observed a significant sex difference between male and female speakers (t-test, p = 0.002), with male proportions falling closer to the 1:1 ratio (see Figure 2, Table 1).
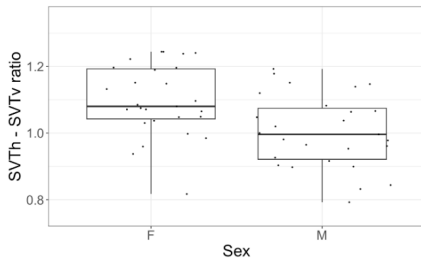

Figure 1. Boxplot of SVT_H–SVT_V ratio for male and female subjects.

Table 1. Summary of SVT_H–SVT_V ratio observed in our sample.

	SVT_H-SVT_V		
	All (N=55)	Male (N=28)	Female (N=27)
Avg.	1.05	1.00	1.10
St.Dev	.12	.12	0.12
Min.	0.79	0.79	0.82
Max.	1.24	1.19	1.24

Discussion

Vocal tracts of extinct hominins

The oft-purported benefits of "roughly equal" vocal tract proportions have long been a staple of research on the evolution of speech (Lieberman, 1991, 2012; Carré et al., 1995; de Boer, 2010a, 2010b; Badin et al., 2014). In particular, Neanderthal phonetic capacities estimated by Lieberman and colleagues – the first modern efforts to attempt a quantification (see overview in Ekström & Moran, 2024) – concluded that the species would have been limited from the extent of fully modern human speech. Lieberman and colleagues argued that the different proportions restricted the ability to create "abrupt midpoint discontinuities" – argued to be necessary to articulate the extreme "quantal" vowels. According to these estimates, and due to a relatively long face and short neck, Neanderthal vocal tracts would have been precluded from vowels [a], [i], and [u], but included [ɪ], [æ], and [ɛ]. While the empirical bases of these estimates (i.e., the assumption that the angle of the basicranium provides reliable information about the shape of vocal tracts) have been refuted (Lieberman & McCarthy, 1999; Fitch & Giedd, 1999), later estimates have effectively reinforced the conclusions of these original efforts. Estimates by McCarthy (reported in Lieberman, 2007) and Barney and colleagues (2012) have both indicated a relatively limited phonetic potential in Neanderthals. On the other hand, Badin et al. (2014) have argued these objections may be overstated. Our data have potential implications for this line of research.

Why proportions matter

The influence on phonetic output resulting from variations in vocal tract anatomy is a topic of substantial interest in speech-centric sciences (Simpson, 2002). For example, Johnson and colleagues (1993) considered consequences resulting from degrees of palatal doming (higher in males); more significant doming effectively results in a greater physical space that must be traversed in order to execute a given phonemic transition that involves that space. More pertinent to our data, de Boer (2010a, p. 351), in modeling the influence of various laryngeal heights found that "there is an optimal larynx height at which the largest

range of signals can be produced and that at this height, the vertical and horizontal parts are approximately equally long." Because sex differences (relatively longer pharynges in males due to a second process of laryngeal descent) emerge during puberty, de Boer (2010) argued that the apparent phonetic efficiency bestowed by a SVT_H-SVT_V configuration which "corresponds closely to human female anatomy", was offset in males for the purpose of "size exaggeration." The format of the present text does not allow for an in-depth review of relevant literature on the influence of variations in vocal tract anatomy on phonetic output; it should be mentioned, however, that Badin et al. (2014) have challenged de Boer's model (but not the results presented by Carré et al., 1995, 2017). Nevertheless, in a broad sense, the influence of vocal tract proportions is likely meaningful. However, works on the phonetic capacities of extinct hominins suffer from an apparent weakness, which must, in light of presently presented data, be adequately addressed.

Toward evolutionary biomechanics

The early estimates by Lieberman and colleagues (1972, p. 297) suggested that [u] was "virtually impossible for the chimpanzee to articulate.". Allegedly, this was due to anatomy which prevented "the tongue body motion found in man." Vocal tract proportions in extant primates is indeed drastically disparate from those observed in humans, and those inferred for extinct hominins; for example, the adult chimpanzee possess an SVT_H more than twice the length of tits SVT_V (Nishimura, 2005). Nevertheless, recent works on the phonetic aspects of great ape call repertoires have documented [u]-like calls, overlapping almost perfectly with vowel formants typically estimated for [u] (Grawunder et al., 2022; Ekström et al., 2023). The articulatory configurations underlying these vowel-like formant dispersions are as of yet unknown. Nonetheless, this evidence suggests that (some) extremes of human vowel space can be achieved by extant non-human great apes by executing different vocal tract configurations than speaking humans (Ekström, 2024a). Moving forward, it will be necessary to recognize that counter to a long-running tradition in "evolutionary phonetics" sciences (overview in Ekström, 2024b), phonetic capacities are not reasonably estimated from isolated vowel production capacities alone. Realistic articulatory modeling will be necessary to determine biomechanics of how vowel tract proportions unlike those of modern humans may (or may not) execute sequences of speech sounds. Toward this end, detailed descriptions of variability in vocal tract anatomy are a pertinent line of evidence.

Acknowledgements

References

Badin, P., Boë, L. J., Sawallis, T. R., & Schwartz, J. L. (2014). Keep the lips to free the larynx: Comments on de Boer's articulatory model (2010). *Journal of Phonetics*, *46*, 161-167.

Barney, A., Martelli, S., Serrurier, A., & Steele, J. (2012). Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 88-102.

Bosma, J. F. (1975). ''Anatomic and physiologic development of the speech apparatus,'' in The Nervous System: Human Communication and its Disorders, edited by D. B. Tower. Raven, New York, Vol. 3, pp. 469–481.

Carré, R., Lindblom, B., & MacNeilage, P.F. (1995). Rôle de l'acoustique dans l'évolution du conduit vocal humain. *Comptes Rendus de l'Académie des Sciences*, Série IIb, 320, 471–476.

Carré, R., Divenyi, P., Mrayati, M. (2017). *Speech: A dynamic process.* De Gruyter.

Crelin, E. S. (1973). *Functional Anatomy of the Newborn*. Yale University Press, New Haven, CT.

de Boer, B. (2010a). Modelling vocal anatomy's significant effect on speech. *Journal of Evolutionary Psychology*, *8*(4), 351–366.

de Boer, B. (2010b). Investigating the acoustic effect of the descended larynx with articulatory models. *Journal of Phonetics*, *38*(4), 679-686.

Ekström, A. G., Moran, S., Sundberg, J., & Lameira, A. R. (2023). PREQUEL: Supervised phonetic approaches to analyses of great ape quasi-vowels. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, 3076–3080, Prague, Czech Republic. https://doi.org/10.31234/osf.io/8aeh4

Ekström, A. G. (2024). Correcting the record: Phonetic potential of primate vocal tracts and the legacy of Philip Lieberman (1934− 2022). *American Journal of Primatology*, e23637.

Ekström, A. G. (2024b). A theory that never was: Wrong way to the "Dawn of speech". *Biolinguistics, 18*, e14285. https://doi.org/10.5964/bioling.14285.

Ekström, A. G. & Moran, S. (2024). What phonetics has to say about Neanderthal (*H. neanderthalensis*) speech capacities. In J. Nölle, L. Raviv, K. E. Graham, S. Hartmann, Y, Jadoul, M. Josserand, T. Matzinger, K. Mudd, M. Pleyer, A. Slonimska, S. Wacewicz, & S. Watson (Eds.), *The Evolution of Language: Proceedings of the 15th International Conference (Evolang XV)*. doi:10.17617/2.3587960.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511–1522.

Grawunder, S., Uomini, N., Samuni, L., Bortolato, T., Girard-Buttoz, C., Wittig, R. M., & Crockford, C. (2022). Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200455.

Kent, R. D. (1981). Articulatory-acoustic perspectives on speech development, in R. Stark (ed.), *Language Behavior in Infancy and Early Childhood*. Elsevier: Amsterdam. (pp. 105–239).

Kent, R. D., and Vorperian, H. K. (1995). Anatomic development of the craniofacial-oral-laryngeal systems: A review. *Journal of Medical Speech-Language Pathology, 3*, 145–190.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, *94*(2), 701-714.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

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*(3884), 1185–1187.

Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*, *74*(3), 287–307.

Lieberman, P. (1991). *Uniquely human: The evolution of speech, thought, and selfless behavior*. Harvard University Press.

Lieberman, P., Laitman, J. T., Reidenberg, J. S., & Gannon, P. J. (1992). The anatomy, physiology, acoustics and perception of speech: essential elements in analysis of the evolution of

human speech. *Journal of Human Evolution*, *23*(6), 447–467.

Lieberman, D. E., & McCarthy, R. C. (1999). The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *Journal of Human Evolution*, *36*(5), 487–517.

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*, *46*(2), 117-128.

Lieberman, P. (2007). The evolution of human speech: Its anatomical and neural bases. *Current Anthropology*, *48*(1), 39-66.

Lieberman, P. (2012). Vocal tract anatomy and the neural bases of talking. *Journal of Phonetics*, *40*(4), 608–622.

Lieberman, P. (2017). Comment on "Monkey vocal tracts are speech-ready". *Science Advances*, *3*(7), e1700442.

Lim, Y., Toutios, A., Bliesener, Y., Tian, Y., Lingala, S. G., Vaz, C., ... & Narayanan, S. S. (2021). A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images. *Scientific Data*, *8*(1), 187.

Negus, V. E. (1949). *Comparative anatomy and physiology of the larynx*. Heinemann.

Nishimura, T. (2005). Developmental changes in the shape of the supralaryngeal vocal tract in chimpanzees. *American Journal of Physical Anthropology*, *126*(2), 193-204.

Simpson, A. P. (2002). Gender-specific articulatory–acoustic relations in vowel sequences. *Journal of Phonetics*, *30*(3), 417-435.

Sorensen, T., Skordilis, Z. I., Toutios, A., Kim, Y. C., Zhu, Y., Kim, J., ... & Narayanan, S. S. (2017, August). Database of Volumetric and Real-Time Vocal Tract MRI for Speech Science. In *Proceedings of Interspeech* 2017 (pp. 645-649).

Stevens, K. N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr & P. B. Denes (Eds.), *Human communication: a unified view* (pp. 51–66). McGrawHill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1-2), 3-45.

Takemoto, H. (2008). Morphological analyses and 3D modeling of the tongue musculature of the chimpanzee (Pan troglodytes). *American Journal of Primatology*, *70*(10), 966–975.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, *117*(1), 338-350.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., ... & Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*, *125*(3), 1666-1678.