Why only a "great" primate discovered speech: A progress report on evolving vocal repertoires

Axel G. Ekström[1†], Margarita Briseño-Jaramillo[2], Carla Pascual-Guàrdia[3], Melissa Berthet[4,5], Kayla Sward[6], Auriane Le Floch[3,7], Franziska Wegdell[4], Alban Lemasson[8,9], Julia Fischer[10], Adrian Soldati[3], Daniel Friedrichs[3,11], Serge Wich[12], Sonja Holmer[13,14], Alexandra B. Bosshard[4], Lauren Halenar-Price[15,16], Adwait Desphande[3,17,18,19,20], Brian M. Shearer[16,21], Steven Moran[3,22], Jens Edlund[1]

Abstract

There are hundreds of extant species of primate. Is it a coincidence that the only known species to develop fluid speech was part of the species-poor primate clade characterized by substantially greater relative body size? In this position paper, we discuss pertinent evidence from four species of American monkey, seven species of Afroeurasian monkey, and multiple species of ape including all extant great ape genera. The evidence indicates that a propensity to utilize distinctive formant dispersions ("vowel-like" qualities) is facilitated by vocal tract length. Smaller primates appear to make greater widespread use of source variations, constrained from effective communicative use of formants by their relatively short vocal tracts and small vocal folds (facilitating high-f_0 signals). Exceptions to these apparent trends are discussed. This work highlights the roles of anatomy in shaping species' vocalizations and vocal behavior and explores several emerging trends of relevance for the evolution of primate call repertoires.

Introduction

This work is a contribution toward the Symposium celebrating the contributions to speech science of Professor Björn Lindblom on his 90th birthday, held at the annual *Fonetik* Congress in Stockholm, Sweden. Lindblom's (1984, 1990 2000) contributions to an evolutionary science of human speech are part of a small but rich literature on human spoken language and communication as an emergent property (e.g., Carré et al., 1995; Studdert-Kennedy, 2005; Ekström & Edlund, 2023). In his 2017 InterSpeech keynote lecture[1], Lindblom explored why humans "gave up" being like the other apes. Here, we explore a

---

[1] youtube.com/watch?v=k1j5wRb4cF0

[1]*Speech, Music & Hearing, KTH Royal Institute of Technology, Sweden;* [2]*Instituto de Neuroetología, Universidad Veracruzana, Mexico;* [3]*Institute of Biology, University of Neuchâtel, Switzerland;* [4]*Department of Evolutionary Anthropology, University of Zürich, Switzerland;* [5]*Center for the Interdisciplinary Study of Language Evolution, University of Zürich, Switzerland;* [6]*Department of Linguistics and Philology, Uppsala University, Sweden;* [7]*Ape Social Mind Lab, Institut des Sciences Cognitives Marc Jeannerod, CNRS, Lyon, France;* [8]*Univ Rennes, Normandie Univ, CNRS, EthoS (Ethologie Animal et Humaine), Rennes, France;* [9]*Institut Universitaire de France, France;* [10]*Cognitive Ethology Lab, German Primate Center, Göttingen, Germany;* [11]*Department of Computational Linguistics, University of Zürich;* [12]*School of Biological and Environmental Sciences, Liverpool John Moores University, UK;* [13]*Lund University Cognitive Science, Department of Philosophy, Lund University, Sweden;* [14]*Faculty of Governance and Global Affairs, Leiden University College The Haag, The Netherlands;* [15]*Department of Biology, Farmingdale State College (SUNY), New York, New York;* [16]*NYCEP Morphometrics Group, American Museum of Natural History, New York, NY, USA;* [17]*Inkawu Vervet Project, Mawana Game Reserve, KwaZulu-Natal, South Africa;* [18]*Department of Collective Behavior, Max Planck Institute of Animal Behavior, Germany;* [19]*Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Germany;* [20]*Max Planck Institute of Animal Behavior, Konstanz , Germany;* [21]*Department of Cell Biology, New York University School of Medicine, USA;* [22]*Department of Anthropology, University of Miami, FL, US.* [†]*Address correspondence to: axeleks@kth.se*

Table 1. Species studied as of May 2024.

Platyrrhines	Black howler monkey
	Geoffroy's spider monkey
	Black-capped squirrel monkey
	Common marmoset
Catarrhines	Vervet monkey
	Lesser spot-nosed monkey
	Sooty mangabey
	Diana monkey
	Thomas' langur
	Guinea baboon
Apes	Gibbon*
	Siamang
Great apes	Bornean orangutan
	Western gorilla
	Western chimpanzee
	Bonobo

*Several species sampled.

similar theoretical issue, asking why, out of all the hundreds of species of primates, the only species to evolve speech, came from a narrow selection of larger bodied taxa. This work combines observations of vocalization data from four species of American monkeys – howler monkeys (*Alouatta pigra*) (Briseño-Jaramillo et al., 2017), spider monkeys (*Ateles geoffroyi*) (Briseño-Jaramillo et al., 2018), Bolivian squirrel monkeys (*Saimiri boliviensis boliviensis*), and common marmosets (*Callithrix jacchus*) – seven species of Old World monkey – including vervet monkeys (*Chlorocebus pygerythrus*) (Deshpande et al., 2023), lesser spot-nosed monkeys (*Cercopithecus petaurista*) (Le Floch et al., 2021), sooty mangabeys (*Cercocebus atys*), Diana monkeys (*Cercopithecus diana*) (Zuberbühler, 2000), Thomas' langurs (*Presbytis thomasi*) (Wich et al., 2003)

and Guinea baboons (*Papio papio*) (Fischer et al., 2002) – as well as apes, including gibbons (*Hylobates* .spp) and siamangs (*Symphalangus syndactylus*), Bornean orangutans (*Pongo pygmaeus*) (Lameira & Wich, 2008; Ekström et al., 2023), Western gorillas (*Gorilla gorilla*) (Salmi et al., 2014) and Western chimpanzees (*Pan troglodytes*) (Grawunder et al., 2022). We outline two main directions and schools of thought that have emerged through this series of comparisons, which will be explored further in subsequent research.

Trends in smaller primates

Chucks, whinnies, and whoops

Across descriptions of primate vocal repertoires, onomatopoeic naming conventions are widespread. For smaller primates, names assigned to species' calls – from the "chucks" of squirrel monkeys, "tcha-kow" calls of lesser spot-nosed monkeys, and the "whinnies" of spider monkeys – often evidently correspond to distinct fundamental frequency variations. The would-be consonant-to-vowel distinction in a "chuck", for example, corresponds to a significant drop from high-frequency phonation or noise (the "ch") to a relatively lower frequency threshold (the "uck"). Such naming conventions stand in contrast with names assigned to calls by larger primates like baboon "wa-hoo", and chimpanzee "hoots", which are named for their apparent and distinct vowel-like qualities, verifiable through spectrographic analysis (Salmi et al., 2014; Boë et al., 2017; Grawunder et al., 2022; Ekström et al., 2023, 2024a).

Acoustics of short vocal tracts

It is a fact of physics that shorter vocal tracts will produce higher resonant frequencies; is this a meaningful determinant of speech-like behavior in primates? In a series of in-prep speech acoustics computer simulation works, we have investigated the impact on

vowel quality of increasingly limiting vocal tract lengths. By scaling vocal tracts down but maintaining vocal tract proportions, we can investigate speech signals corresponding to a hypothetical "small" speaker. These simulations suggest that length of the vocal tract (and therefore the animal) may be a determinant as to whether an animal develops a "spectral peak-based" vocal repertoire (i.e., one where formant-like patterns distinguish some or all contrasting vocal signals). When vocal tract lengths are shorter, even highly distinct formant dispersions, such as those observed for close front unrounded vowel [i], become markedly more difficult to distinguish, owing to higher formants being shifted up and toward the higher end of the spectrum of human perceivability (Table 2). For short vocal tracts, the lowest formant frequency (F1) may be the greatest determinant of vowel quality. Speech, as spoken by all normally developing humans, requires a long enough vocal tract.

Trends in larger primates: Vowel quality and vowel-like quality

Vowel quality is traditionally reducible to the constellation of the first and second formants (F2), where F1 roughly corresponds to jaw height and degree of stricture, and F2 corresponds to the position of the stricture (Fant, 1960). The third formant (F3) is traditionally said to correspond to the degree of mouth closure. These assumptions, while simplistic, allow for fairly reliable inference of vocal tract dynamics underlying speech behavior in real time, even from acoustic data alone. However, reflecting the comparative shapes of nonhuman vocal tracts (Negus, 1949; Harrison, 1995), these assumptions cannot be extended to primates a priori. Formant data has been reported for baboons (Fischer et al., 2002; Boë et al., 2017), orangutans (Ekström et al., 2023), gorillas (Ekström et al., 2024a),

and chimpanzees (Grawunder et al., 2022). In chimpanzees, these support consonant-vowel spectrographic patterns (Ekström et al., 2024b). There are also indications of similar speech-like properties in vocalizations by howler monkeys (Schön Ybarra, 1986) - a markedly smaller primate (relative to great apes) and American monkey. These studies have, however, been performed with formants averaged over a window of time: reported data are effectively snapshots, which cannot be used to infer vocal tract dynamics per se. To what extent do these "vowel-like" vocalizations reflect human-like production?

Exceptions to the trends?

A highly derived anatomy

We do not mean to suggest the tendencies discussed above are universal or without exceptions. With regard to relatively "small" primates, for example, the vocal anatomy of the Alouatta genus is likely one of the most derived in nature (Figure 1), likely reflecting a unique eco-behavioral strategy (Dunn et al., 2015; Youlatos et al., 2015). Effectively, the presence of additional resonant frequencies at relatively low

Table 2. Formants for [i], scaled to the average adult female VTL, and to VTL = 8 cm. Scaled linearly. Original values (17 cm) for Swedish male speakers from Fant (1959). Maintaining consistent vocal tract proportions, shorter vocal tracts will produce formants increasingly beyond the optimal window for perception, and their contribution to vowel quality becomes increasingly small.

VTL	F1	F2	F3
8	574	4866	6396
14	328	2780	3655
17	270	2290	3010

thresholds induces a "vowel-like" quality in a manner distinct from that of other primates (large or otherwise). Howlers have achieved this vocal repertoire via the evolution of a highly specialized hyolaryngeal complex, typified by a hypertrophied hyoid bulla, that is theorized to act as a resonating chamber which aids the production of their ~90 dB vocalizations. The divergence in aerodigestive tract anatomy from the apes, which exhibit larger body mass and utilize soft-tissue air sacs to aid in vocalization, demonstrates a possible case of parallel evolution towards the ability to produce "vowel-like" vocalizations. The howler monkeys' case underscores vocal anatomy's relevance to the evolution of vocal production strategies.

Figure 1. Bisected Howler monkey (Alouatta seniculus) vocal tract. The prominent species-typical hyoid bulla is visible inferior to the mandible. The scale bar is 1 cm. Photo credit: B. Shearer.

High-frequency calls in a large primate

Bonobos are great apes of the *Pan* genus. Their vocal repertoire is mainly composed of relatively high-frequency calls (Hohmann & Fruth, 1994): there is no reported instance of any bonobo producing a call overlapping in vowel-like quality to that of, for example, orangutan long calls (Ekström et al., 2023). Bonobos possess specialized vocal fold anatomy, including markedly shortened vocal folds (Grawunder et al., 2018). Rather than a counterexample, the case of bonobos likely provides another example of systems of vocalizations reflecting complex combinations of selection pressures.

Discussion

The vocal tracts of subadult and adult chimpanzees overlap in length with those of humans (Nishimura, 2005); while those of adult baboons overlap with human preteens (Boë et al., 2017). Both allow for a range of vowel-like qualities (Boë et al., 2017; Berthommier, 2020; Grawunder et al., 2022). Here, a necessary distinction must be made. Namely, Boë et al. (2017) interpret the argument that the reconfiguration of the hominin vocal tract (involving a marked expansion of the pharynx and descent of the tongue root) expanded the phonetic potential of human ancestors to mean that all speech would be impossible without it (Boë et al., 2017). This interpretation is refuted in Ekström (2024a).

In reality, limitations on primate phonetic potential are not seriously disputed by any speech-centric work (Ekström, 2024b) and become more untenable still in light of speech biomechanics (Takemoto, 2008; Iwasaki et al., 2019; Ekström & Edlund, 2023). Analyses of primate lingual biomechanics (Takemoto, 2001, 2008) demonstrate that "flat" primate tongues (Negus, 1949) possess disparate degrees of freedom, meaning that any fluid speech capacity available to primates would be markedly different from that of humans. Literature on the evolution of speech has concentrated on a hypothetical conflict between "anatomical" and "neural" evolution (Lieberman et al., 1972; Fitch et al., 2016). Accumulated evidence shows that neither sufficiently explains the differences between primate communication systems and human spoken language (Berthommier, 2020;

Ekström, 2024b). It is imperative that we move toward an evolutionarily coherent model.

There are several hundreds of extant species of primate - an order characterized by relatively large brains relative to body size, a pronounced selection pressure for visual acuity, and group social complexity. The vocalizations of primates are highly diverse, and human spoken language constitutes a particularly remarkable example. We have expounded on a range of observations from primate vocal behavior, with two main takeaways.

First, there are likely ecological, social, and morphological underpinnings that determine the evolution of primate vocal repertoires, which may be studied with reference to comparative acoustic data. Second, one such principle underlying the structural and acoustic bias of human speech communication was likely permissible through the size of our ancestors. The characteristics of human speech were unlikely to evolve in a small primate. Lindblom's (1984, 2000) contributions to speech-centric science explore the foundational principles by which speech behaviors emerge. Here, we have sought to outline avenues for future research by which we hope to and understand the organizing principles structuring the development of primate vocal repertoires.

Acknowledgments

References

Berthommier, F. (2020, March). Monkey vocal tracts are not so "speech ready". In N. H. Bernadoni & L. Bailly (Eds.), *Proceedings of 12th International Conference on Voice Physiology and Biomechanics*.

Boë, L. J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., ... & Fagot, J. (2017). Evidence of a vocalic proto-system in the baboon (Papio papio) suggests pre-hominin speech precursors. *PloS One*, *12*(1), e0169321.

Briseño-Jaramillo, M., Biquand, V., Estrada, A., & Lemasson, A. (2017). Vocal repertoire of free-ranging black howler monkeys (*Alouatta pigra*): call types, contexts, and sex-related contributions. *American Journal of Primatology*, *79*(5), e22630.

Briseno-Jaramillo, M., Ramos-Fernandez, G., Palacios-Romo, T. M., Sosa-López, J. R., & Lemasson, A. (2018). Age and social affinity effects on contact call interactions in free-ranging spider monkeys. *Behavioral Ecology and Sociobiology*, *72*, 1-17.

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..

Deshpande, A., van de Waal, E., & Zuberbühler, K. (2023). Context-dependent alarm responses in wild vervet monkeys. *Animal Cognition*, *26*(4), 1199-1208.

Dunn, J. C., Halenar, L. B., Davies, T. G., Cristobal-Azkarate, J., Reby, D., Sykes, D., ... & Knapp, L. A. (2015). Evolutionary trade-off between vocal tract and testes dimensions in howler monkeys. *Current Biology*, *25*(21), 2839-2844.

Ekström, A. G., & Edlund, J. (2023). Evolution of the human tongue and emergence of speech biomechanics. *Frontiers in Psychology*, *14*, 1150778. https://doi.org/10.3389/fpsyg.2023.1150778

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 2023* (pp. 3076–3080). Guarant: Prague. https://doi.org/10.31234/osf.io/8aeh4

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

Ekström, A. G. (2024b). 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., Nellissen, L., Bortolato, T., Crockford, C., Edlund, J., Zuberbühler, K., ... Moran., S., (2024a). Phonetic properties of chimpanzee, gorilla, and orangutan hoots tell a uniform story and point to new frontiers. In Nölle, J., Raviv, L., Graham, K. E., Hartmann, S., Jadoul, Y., Josserand, M., Matzinger, T., Mudd, K., Pleyer, M., Slonimska, A., Wacewicz, S. & Watson, S. (Eds.): *The Evolution of Language: Proceedings of the 15th International Conference (Evolang XV)*. doi:10.17617/2.3587960.

Ekström, A. G., Gannon, C., Edlund, J., Moran, S., & Lameira, A. (2024b). No missing link for chimpanzee speech production. https://doi.org/10.31234/osf.io/359bd

Fant, G. (1959). Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Technics No. 1.*

Fant, G. (1960). The acoustic theory of speech production. Mouton: The Hague.

Fischer, J., Hammerschmidt, K., Cheney, D. L., & Seyfarth, R. M. (2002). Acoustic features of male baboon loud calls: Influences of context, age, and individuality. *The Journal of the Acoustical Society of America*, *111*(3), 1465-1474.

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*(12), e1600723.

Grawunder, S., Crockford, C., Clay, Z., Kalan, A. K., Stevens, J. M., Stoessel, A., & Hohmann, G. (2018). Higher fundamental frequency in bonobos is explained by larynx morphology. *Current Biology*, *28*(20), R1188-R1189.

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.

Harrison, D. F. N. (1995). *Comparative anatomy and physiology of the mammalian larynx*. Cambridge University Press.

Hohmann, G., & Fruth, B. (1994). Structure and use of distance calls in wild bonobos (*Pan paniscus*). *International Journal of Primatology*, *15*, 767-782.

Iwasaki, S. I., Yoshimura, K., Shindo, J., & Kageyama, I. (2019). Comparative morphology of the primate tongue. *Annals of Anatomy-Anatomischer Anzeiger*, *223*, 19-31.

Lameira, A. R., & Wich, S. A. (2008). Orangutan long call degradation and individuality over distance: a playback approach. *International Journal of Primatology*, *29*, 615-625.

Le Floch, A., Bouchard, A., Gallot, Q., & Zuberbühler, K. (2021). Lesser spot-nosed monkeys coordinate alarm call production with associated Campbell's monkeys. *Behavioral Ecology and Sociobiology*, *75*, 1-14.

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

Lindblom, B. (1984). Can the models of evolutionary biology be applied to phonetic problems? *Proceedings of the Xth International Congress of Phonetic Sciences* (pp. 67-81).

Lindblom, B. (1990). On the notion of "possible speech sound". *Journal of Phonetics*, *18*(2), 135-152.

Lindblom, B. (2000). Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica*, *57*(2-4), 297-314.

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.

Salmi, R., & Doran-Sheehy, D. M. (2014). The function of loud calls (Hoot Series) in wild western gorillas (*Gorilla gorilla*). *American Journal of Physical Anthropology*, *155*(3), 379-391.

Schön Ybarra, M. A. (1986). Loud calls of adult male red howling monkeys (*Alouatta seniculus*). *Folia Primatologica*, *47*(4), 204-216.

Studdert-Kennedy, M. (2005). How did language go discrete. In M. Tallerman (Ed.), *Language origins: Perspectives on evolution* (pp. 48-67). Oxford University Press.

Takemoto, H. (2001). Morphological analyses of the human tongue musculature for three-dimensional modeling. Journal of Speech, Language, and Hearing Research, *44*(1).

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.

Wich, S. A., Koski, S., de Vries, H., & van Schaik, C. P. (2003). Individual and contextual variation in Thomas langur male loud calls. *Ethology*, *109*(1), 1-13.

Youlatos, D., Couette, S., & Halenar, L. B. (2015). Morphology of howler monkeys: A review and quantitative analyses. In M. M. Kowalewski, P. A. Garber, L. Cortés-Ortiz, B. Urbani, & D. Youlatos (Eds.), *Howler monkeys: Adaptive radiation, systematics, and morphology* (pp. 133-176). Springer.

Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal Behaviour, 59*(5), 917-927.