

EVOLUTIONARY BIOLOGY

Response to Lieberman on “Monkey vocal tracts are speech-ready”

W. Tecumseh Fitch,^{1,2*} Bart de Boer,³ Neil Mathur,^{4,5} Asif A. Ghazanfar^{4,5,6}

We thank P. Lieberman for his technical comment, and we are pleased that he accepts our data, methods, and results and agrees with our main conclusion: that a macaque’s vocal tract would be able to produce speech sounds if macaques had the required neural control. However, we cannot agree that our findings, which expand the phonetic potential of macaques eightfold relative to that reported in his seminal 1969 paper, in any sense constitute a “replication” of that study or demonstrate the correctness of his earlier conclusions.

To recap, both studies used measurements of macaque monkey vocal tracts to create a computer model, which was then queried to determine what vocalizations it could potentially produce: a space representing the “phonetic potential” of that vocal tract. The key difference between the two studies is that our vocal tract measurements were derived from x-rays of living monkeys vocalizing and communicating (1), whereas the measurements of Lieberman *et al.* [(2), p. 1186] were derived from a single cast of a dead monkey, with possible perturbations “estimated” by “manipulating ... an anesthetized monkey.” We believe that this difference in the quality of the input data is responsible for the key difference in our results: an eightfold increase in the macaque phonetic potential as estimated by our model [see our Fig. 3 in (1)]. Going beyond Lieberman’s original study, we also generated five “monkey vowels” that optimally partitioned this enlarged acoustic space. Perceptual experiments then showed that humans readily discriminate between these five vowels. Five vowels were chosen because that is the modal number of vowels in human languages around the world, although the specific vowels vary, of course, from language to language (3, 4). Given that nonhuman primate formant perception is very similar to that of humans (5, 6), this finding indicates that a primate vocal tract could easily produce enough vowels to support a large communicative vocabulary if the neural capabilities to exploit this phonetic range were present.

We thus concluded that vocal anatomy cannot be the key causal factor explaining the complete lack of speech in nonhuman primates. That vocal tract anatomy is the crucial factor is widely believed. For some textbook examples, “Early experiments to teach chimpanzees to communicate with their voices failed because of the insufficiencies of the animals’ vocal organs” [(7), p. 402]. “The vocal tracts of chimps are physiologically unsuited to producing speech, and this difference alone could account for their lack of progress” [(8), p. 59]. “Chimpanzees have a vocal tract that makes speech production essentially impossible” [(9), p. 75]. These quotations, which could readily be multiplied, illustrate a very widespread misconception that nonhuman

primate vocal tracts are inadequate to produce speech irrespective of neural control. This myth was the primary target of our empirical study. Our hope is that this pervasive belief is finally laid to rest by our study, and we are pleased that Lieberman questions neither the methods or data we used nor our conclusion that neural factors are crucial.

Our only debate with Lieberman thus concerns the relative importance of neural control versus vocal anatomy in explaining human speech abilities. We can hardly disagree with his statement that speech entails “both anatomy and brains,” but the question we addressed is whether changes in anatomy were necessary for speech capacities to evolve; we conclude that, because macaque anatomy could produce a wide range of discriminable vowels, they were not. This does not mean that human vocal anatomy has not changed—it has (more below)—but these anatomical changes were neither necessary nor sufficient for the evolution of spoken language.

THE MYTHICAL POWER OF [i]

A central issue for Lieberman is the ability in humans and the putative lack thereof in monkeys to produce certain extreme vowels, which Lieberman terms “quantal vowels,” for example, [i], [a], and [u] (as in “beet,” “cot,” and “root”), which are common but by no means universal in human languages.

First, contra Lieberman, we did not show that “the monkey vocal tract cannot produce” these vowels—only that we never observed monkeys doing so. Our approach, relying solely on configurations actually observed in communicating monkeys, is by its nature very conservative and cannot support any strong claims about inabilities. Thus, it remains possible that macaque vocal anatomy could support these extreme vowels, if they were important in macaque communication. Furthermore, our monkey model can produce a low vowel very close to the human “quantal vowel” [a].

Second, we are skeptical that [i] (or any other vowel) plays a necessary role in spoken language. Any vocal tract, of any species, has extreme configurations, and whatever formant patterns are produced by those configurations can play the same “perceptual anchoring” role that Lieberman posits for the quantal vowels in speech. Furthermore, the quantal nature of vocal production explored by Stevens (10–12) in his seminal theoretical studies is not limited to the human vocal tract: Any vocal tract will have regions where small articulatory changes have large acoustic consequences and vice versa, thus rendering certain vocal tract configurations “quantal.” Stevens never suggested that there is anything special about humans in this respect.

Third, regarding vocal tract normalization, there is now abundant evidence that nonhuman animals can use formant patterns to estimate vocal tract length (and from it, body size) from a variety of roars, grunts, and bellows (13–16). Humans can also readily gauge vocal tract

Copyright © 2017
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Department of Cognitive Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. ²Haidhof Research Station, University of Vienna/University of Veterinary Medicine Vienna, Bad Vöslau, Austria. ³Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. ⁴Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA. ⁵Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. ⁶Department of Psychology, Princeton University, Princeton, NJ 08544, USA. *Corresponding author. Email: tecumseh.fitch@univie.ac.at

length from vowels other than [i] (17, 18). Thus, there is no evidence that [i] is necessary for vocal tract normalization to occur. Furthermore, although formants 1 and 2 play a key role in discriminating between vowels, in real speech stimuli, higher formants (formants 3 to 5) are always present and provide a clear indication of overall vocal tract length, independent of which particular vowel is being produced. Together, these data suggest that vocal tract normalization, via formants, is a basic and widely shared capability among mammals and does not require special vowels or specialized vocal morphology.

Of course, we do not claim that “monkey speech” (Lieberman’s term) would be identical to human speech—both common sense and the results of our study dictate that they would differ (as listening to our simulated examples indicates). However, this is irrelevant to the central evolutionary issue: Would monkey speech (or “chimpanzee speech,” “Neandertal speech,” etc.) have an adequate number of acoustically discriminable phonemes to support a large vocabulary of discriminable words? Given the greatly enlarged macaque phonetic space we found in our study, relative to Lieberman’s original estimates, we concluded that the answer is yes. Whether the vowels are identical to those of human languages is not the issue, any more than differences in the vowel systems of Spanish, English, Danish, or Arernte (which has a reduced “vertical” vowel space) affect their overall suitability for linguistic communication.

MODELS OF NEANDERTAL VOCAL CAPABILITIES

Lieberman makes much of differences between our 2016 study and de Boer and Fitch’s earlier critique of Boë and colleagues’ models of speech evolution (19). This earlier paper focused on attempts to model the speech capabilities of extinct hominids, particularly Neandertals, a topic of perennial dispute. However, this long-running and perhaps unresolvable debate is tangential to our 2016 paper, which concerns empirical data derived from living primates, and not speculation concerning extinct species.

Our critique of Boë and colleagues’ studies (which are themselves critical of Lieberman’s work) discussed what we believe to be flaws in their modeling approach (20, 21). Briefly, shape parameters derived from human vocal tracts play a central role in those models, meaning that they would find any vocal tract, regardless of its anatomy, able to make the same vowel range. Unfortunately, accepting our critique of Boë’s studies does not render Lieberman’s models or conclusions correct: They too are flawed, because of inadequate input data, as our new paper shows. We find that neither Boë and colleagues’ strong positive arguments nor Lieberman’s negative ones are compelling.

We hesitate to be drawn into speculation about Neandertal speech capabilities, especially because we are critical of attempts to reconstruct speech capabilities based on fossils (19, 22). We would not expect Neandertal capabilities to be less than those of macaques. Beyond that, the most solid data available are those on thoracic canal size [which was larger in Neandertals than earlier hominids, suggesting improved breath control (23)] and the loss of laryngeal air sacs. Air sacs are present in all great apes and so, by inference, in our common ancestor with chimpanzees; a fossil australopithecine hyoid bone strongly suggests that air sacs were present in these fossil hominids (24, 25). Because Neandertals’ hyoid anatomy matches our own (26), these air sacs were probably already lost in that species, as in our own. This is the clearest indication, we think, that vocal anatomy has changed significantly during human evolution, and simulations have suggested that the loss of air sacs would stabilize speech sounds (27).

CONCLUSION

In conclusion, we did not (and do not) claim that monkey speech would sound precisely like human speech, only that a monkey vocal tract would be able to produce clearly intelligible speech. We agree with Lieberman that “human vocal anatomy played a role in the evolution of speech”—for example, the loss of laryngeal air sacs or the descent of the human larynx presumably changed the acoustic details of the specific speech sounds that we make. However, this does not mean that those changes played a key causal role in our ability to speak, or the inability of nonhuman primates to imitate human speech (or produce similar complex vocalizations of their own). For that, changes in the neural circuitry for speech control were necessary (28–30).

Our paper thus reached the same conclusion as that reached by Darwin a century ago, that “as the voice was used more and more, the vocal organs would have been strengthened and perfected ... But the relation between the continued use of language and the development of the brain, has no doubt been far more important.” (31). This conclusion seems to us to be strongly favored by all available data, particularly our new study and other recent work (32), and Lieberman’s technical comment provides no new grounds for disputing it.

REFERENCES

1. W. T. Fitch, B. de Boer, N. Mathur, A. A. Ghazanfar, Monkey vocal tracts are speech-ready. *Sci. Adv.* **2**, e1600723 (2016).
2. P. H. Lieberman, D. H. Klatt, W. H. Wilson, Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* **164**, 1185–1187 (1969).
3. B. de Boer, *The Origins of Vowel Systems* (Oxford Univ. Press, 2001).
4. I. Maddieson, *Patterns of Sounds* (Cambridge Univ. Press, 1984).
5. W. T. Fitch, J. B. Fritz, Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J. Acoust. Soc. Am.* **120**, 2132–2141 (2006).
6. R. D. Hienz, A. M. Jones, E. M. Weerts, The discrimination of baboon grunt calls and human vowel sounds by baboons. *J. Acoust. Soc. Am.* **116**, 1692–1697 (2004).
7. D. Crystal, *The Cambridge Encyclopedia of Language* (Cambridge Univ. Press, ed. 2, 2003).
8. T. Harley, *The Psychology of Language: From Data to Theory* (Psychology Press, ed. 4, 2014).
9. E. Hoff, *Language Development* (Wadsworth, 2008).
10. K. N. Stevens, Quantal nature of speech, in *Human Communication: A Unified View*, E. E. David, P. B. Denes, Eds. (McGraw Hill Publishers, 1972).
11. K. N. Stevens, On the quantal nature of speech. *J. Phonetics* **17**, 3–46 (1989).
12. K. N. Stevens, *Acoustic Phonetics* (MIT Press, 1998).
13. B. D. Charlton, W. A. H. Ellis, J. Brumm, K. Nilsson, W. T. Fitch, Female koalas prefer bellows in which lower formants indicate larger males. *Anim. Behav.* **84**, 1565–1571 (2012).
14. W. T. Fitch, J. P. Kelley, Perception of vocal tract resonances by whooping cranes *Grus americana*. *Ethology* **106**, 559–574 (2000).
15. A. A. Ghazanfar, H. K. Turesson, J. X. Maier, R. van Dinther, R. D. Patterson, N. K. Logothetis, Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr. Biol.* **17**, 425–430 (2007).
16. D. Reby, K. McComb, B. Cargnelutti, C. Darwin, W. T. Fitch, T. Clutton-Brock, Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proc. R. Soc. B* **272**, 941–947 (2005).
17. K. Pisanski, P. J. Fraccaro, C. C. Tighe, J. J. O’Connor, S. Röder, P. W. Andrews, B. Fink, L. M. DeBruine, B. C. Jones, D. R. Feinberg, Vocal indicators of body size in men and women: A meta-analysis. *Anim. Behav.* **95**, 89–99 (2014).
18. K. Pisanski, D. Rendall, The prioritization of voice fundamental frequency or formants in listeners’ assessments of speaker size, masculinity, and attractiveness. *J. Acoust. Soc. Am.* **129**, 2201–2212 (2011).
19. B. de Boer, W. T. Fitch, Computer models of vocal tract evolution: An overview and critique. *Adapt. Behav.* **18**, 36–47 (2010).
20. L.-J. Boë, J.-L. Heim, K. Honda, S. Maeda, The potential Neandertal vowel space was as large as that of modern humans. *J. Phonetics* **30**, 465–484 (2002).
21. L.-J. Boë, S. Maeda, J.-L. Heim, Neandertal man was not morphologically handicapped for speech. *Evol. Commun.* **3**, 49–77 (1999).
22. W. T. Fitch, Fossil cues to the evolution of speech, in *The Cradle of Language*, R. P. Botha, C. Knight, Eds. (Oxford Univ. Press, 2009), pp. 112–134.
23. A. M. MacLarnon, G. P. Hewitt, The evolution of human speech: The role of enhanced breathing control. *Am. J. Phys. Anthropol.* **109**, 341–363 (1999).

24. Z. Alemseged, F. Spoor, W. H. Kimbel, R. Bobe, D. Geraads, D. Reed, J. G. Wynn, A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature* **443**, 296–301 (2006).
25. W. T. Fitch, *The Evolution of Language* (Cambridge Univ. Press, 2010).
26. B. Arensburg, L. A. Schepartz, A. M. Tillier, B. Vandermeersch, Y. Rak, A reappraisal of the anatomical basis for speech in Middle Palaeolithic hominids. *Am. J. Phys. Anthropol.* **83**, 137–146 (1990).
27. B. de Boer, Acoustic analysis of primate air sacs and their effect on vocalization. *J. Acoust. Soc. Am.* **126**, 3329–3343 (2009).
28. W. T. Fitch, Empirical approaches to the study of language evolution. *Psychon. Bull. Rev.* **24**, 3–33 (2017).
29. U. Jürgens, Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* **26**, 235–258 (2002).
30. G. F. Striedter, *Principles of Brain Evolution* (Sinauer, 2004).
31. C. Darwin, *The Descent of Man and Selection in Relation to Sex* (John Murray, ed. 1, 1871).
32. L.-J. Boë, F. Berthommier, T. Legou, G. Captier, C. Kemp, T. R. Sawallis, Y. Becker, A. Rey, J. Fagot, Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests pre-hominin speech precursors. *PLOS ONE* **12**, e0169321 (2017).

Acknowledgments

Author contributions: W.T.F. and B.d.B. wrote the manuscript; all authors edited and approved the final paper. **Competing interests:** The authors declare that they have no competing interests.

Submitted 1 June 2017

Accepted 1 June 2017

Published 7 July 2017

10.1126/sciadv.1701859

Citation: W. T. Fitch, B. de Boer, N. Mathur, A. A. Ghazanfar, Response to Lieberman on “Monkey vocal tracts are speech-ready.” *Sci. Adv.* **3**, e1701859 (2017).