Yawning at the dawn of speech: A closer look at monkey formant space

Caleb Everett
caleb@miami.edu

Fitch and colleagues have offered a methodologically innovative and headline-garnering study that shines new light on the speech capacities of monkey vocal tracts, presenting evidence against the hypothesis that humans' production of peripheral vowels was crucial to the evolution of speech. (1) Along with the elucidative data presented in their study, however, the authors offer some strong conclusions that, for the moment, remain unwarranted. They claim, for instance, that the macaque vocal tract could "easily produce an adequate range of speech sounds." This degree of conviction is, for reasons that should crystallize in this response, surprising.

The shape of the projected monkey formant space varies markedly in accordance with the assumptions made regarding the vocal tract configurations that could be naturally deployed during speech production. For instance, much of the modeled formant space in (1), crucial to the entire enterprise, owes itself to three outlier vocal tract configurations. These configurations are denoted by the rightmost dots in Figure 2 of (1) and, we are told, represent instances in which a macaque was yawning. The straightforward inclusion of these three configurations is somewhat problematic. The configurations reflect maximal/near-maximal mandibular lowering. "Low" vowels in human speech are not actually characterized by such pronounced oral aperture. While it is possible for people to phonate with this vocal tract configuration, it seems telling that no human language has incorporated such an articulatory configuration into its phonemic inventory. Humans produce low vowels (with high first formant [F1] values) with much less movement of the mandible. In fact, we usually only lower our jaws about 1 cm during speech, only a fraction of the extent of possible lowering. (2) Judging from Fitch et al.'s own findings, macaques are not capable of producing high F1 values without extremely pronounced jaw lowering. The authors claim to "definitively" demonstrate that monkeys have speech-ready vocal tracts. But if they conclude that monkey vocal tracts exhibit speech-readiness, they should also emphasize that human vocal tracts exhibit speech-readiness of a completely different magnitude. In short, these new findings may shift the debate about vocal-tract speech-readiness, but they certainly do not resolve it. (It is worth noting that the presumed import of this readiness by some scholars is itself debatable. After all, other animals do not necessarily need peripheral vowels/human-sounding speech in order to have some form of language or even speech.)

If we exclude the three aforementioned data points, the shape of the potential monkey formant space is crucially altered since it then excludes peripheral vowels with high F1 values. Such vowels are ubiquitous in phonemic inventories across the world's thousands of languages. Furthermore, the authors of (1) acknowledge that their estimate of macaque formant space does not allow for a high front vowel (e.g. the vowel sound in *keep*). And, as they note, this vowel also plays a key role in speech. It is worth stressing that no vowel inventories in the world's languages lack both low vowels and high front vowels. In short, it is still debatable if, or

at least how naturally, monkeys can produce peripheral vowels that are crucial to all contemporary spoken languages and, to some researchers anyhow, may well have played a key role at the dawn of speech.

Despite the authors' claim that their approach offers a "highly conservative estimate of potential acoustic output," in fact a more conservative approach to the data would have been to place greater emphasis on those macaque vocal configurations that are analogous to configurations observed during normal human speech (even if considering all configurations is also worthwhile). In Figure A, I offer a new depiction of monkeys' potential F1 and F2 values. This depiction is based on the vocal tract configurations in (1) that were actually produced with vocalizations. As can be seen in the figure, the formant space falls somewhere in between the size of the formant space suggested in Figure 2 of (1) and that based on monkey cadaver analysis. (3) The contrast between the various formant spaces at play can be made more precise by calculating their size. One approach to doing so is offered in Figure A. (The area of an irregular polygon representing a particular formant space is calculated by using the peripheral F1-F2 pairs as the x-y coordinates of the polygon vertices.) Comparing the formant spaces in this way, we see that the macaque vowel space actually evident during vocalizations is 3.2×10^5 Hz^2, about 37% the size of a human female's during speech (4), which is about 8.7×10^5 Hz^2. The size of the macaque formant space described in (1) is approximately 5.6×10^5 Hz^2. Even that latter figure, potentially inflated by yawning configurations, is significantly smaller than the human female's formant space. If we consider all macaque vocal tract configurations but the three with pronounced mandibular lowering, the area of the formant space is 4.6×10^5 Hz^2 and, crucially, lacks both low and high-front vowel configurations. (The F1-F2 pairs used for these formant-space calculations are available from the author.) It should be stressed that, regardless of the assumptions made, the modeled macaque formant space exceeds that based on cadaver analysis (3). This is an important finding, to be sure, but perhaps greater caution should be taken in assessing its implications.
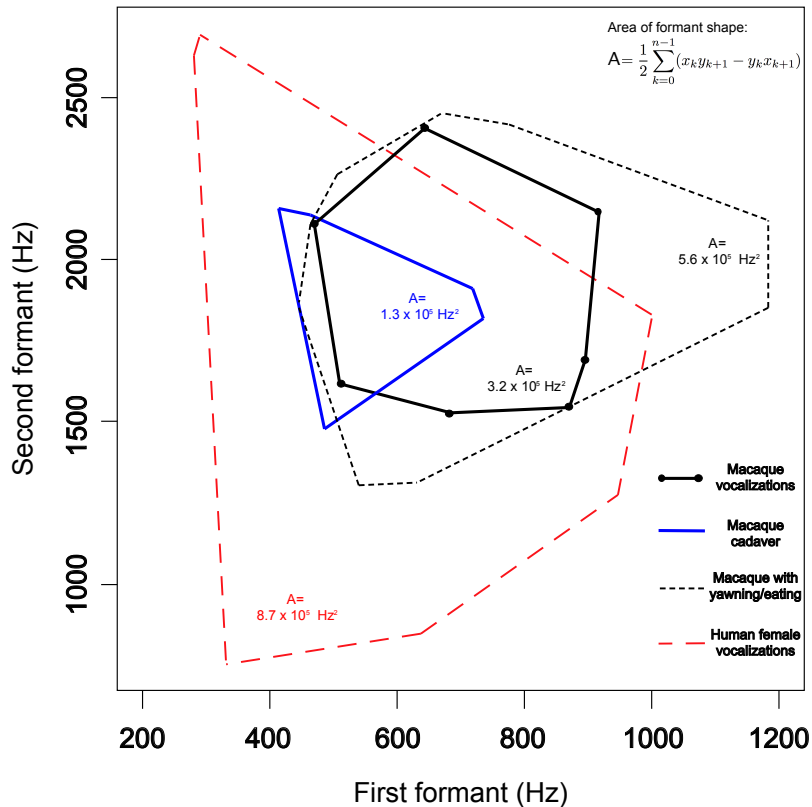
FIGURE A. Comparison of modeled formant spaces.

The projected macaque formant space presented in (1) is potentially problematic for another reason as well. The formant space in Figure 2 of (1) is based on 11 peripheral vocal tract configurations (including the three yawning ones), out of 99 recorded configurations. Most of these 99 configurations are actually much more centrally and densely distributed. In contrast, human peripheral vowels are well representative of speech and, again, they require little mandibular movement. (4) This cross-species contrast belies the simplicity of the conclusions in (1).

Additionally, given the strength of their claims, it would have been useful if Fitch and colleagues had contrasted their modeled monkey formant space with the modeled formant space of a human, i.e. contrasted formant spaces generated via the same methods. One suspects that, if a modeled human formant space was also based on extreme vocal tract configurations (like those used in yawning or screaming), it would be larger than the actual formant space naturally produced in speech. At the least, such a contrastive approach might have better contextualized the authors' calculations of monkey formants.

Finally, it is worth stressing that the account developed in (1) is focused entirely on speech production, overlooking perception (though this point could be made of much of the "neural" versus "peripheral" debate). The authors conclude that, because monkeys putatively have a speech-ready vocal tract, they must not have a speech-ready brain. While many specialists may

agree that monkeys lack a speech-ready brain, this conclusion is somewhat of a non-sequitur in this particular context as it ignores the fact that monkeys and humans differ with respect to their hearing apparati. After all, the human cochlea is particularly well adapted for sounds in the range of 2-5 kHz, a range in which many vowel formants distribute. Peak hearing sensitivity differs somewhat for monkeys and other primates, and monkey hearing is much more sensitive to higher frequencies (as high as 45 kHz, while human hearing is generally not sensitive to frequencies exceeding 20 kHz). (5) I am not claiming that monkeys cannot discriminate vowels, but am suggesting that the extent to which such cross-species cochlear disparities impact acuity of formant discrimination, and potentially speech evolution, is unresolved.

Given the turbid waters we are swimming in here, it seems premature to conclude that the obstacles monkeys confront vis-à-vis language evolution can be confidently categorized as cerebral in etiology. Fitch and colleagues have produced a study with fascinating data, but the conclusions they have drawn from those data are too simple. The age-old question at the heart of their inquiry, viz. whether humans' capacity for speech is due to anatomical or mental factors, or some combination thereof, is not yet "definitively" answered.

References

1. W.T. Fitch, B. d Boer, N. Marthur, A.A. Ghazanfar, Monkey vocal tracts are speech-ready. *Sci. Adv.* 2, e1600723 (2016).

2. K.N. Stevens. Acoustic Phonetics. MIT Press (2000), p. 24.

3. 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).

4. G.E. Peterson, H. L. Barney, Control methods used in a study of the vowels. *J. Acoust.Soc. Am. 24*, 175–184 (1952).

5. W.C. Stebbins, *Animal psychophysics: The design and conduct of sensory experiments*. Springer US (2013), p. 293.