

# Phonetic properties of chimpanzee, gorilla, and orangutan hoots tell a uniform story and point to new frontiers

Axel G. Ekström<sup>1\*</sup>, Lara Nellissen<sup>2,3</sup>, Tatiana Bortolato<sup>4,5,6</sup>, Catherine Crockford<sup>4,5,6</sup>, Jens Edlund<sup>1</sup>, Shelly Masi<sup>3</sup>, Klaus Zuberbühler<sup>2</sup>, Adriano R. Lameira<sup>7</sup>, Roman W. Wittig<sup>4,5,6</sup>, Sven Grawunder<sup>4,8</sup>, and Steven Moran<sup>2,9</sup>

\*Corresponding Author: axeleks@kth.se

<sup>1</sup>Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

<sup>2</sup>Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

<sup>3</sup>Ecoanthropologie (EA), Centre National de la Recherche Scientifique/Museum National d'Histoire Naturelle, University Paris Diderot, Sorbonne Paris Cite, Musée de l'Homme, Paris, France.

<sup>4</sup>Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

<sup>5</sup>Ape Social Mind Lab, Institut des Sciences Cognitives, Lyon, France

<sup>6</sup>Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques, Ivory Coast

<sup>7</sup>Department of Psychology, University of Warwick, Coventry, UK

<sup>8</sup>Department of Speech Science and Phonetics, University of Halle–Wittenberg, Halle, Germany

<sup>9</sup>Department of Anthropology, University of Miami, Coral Gables, FL, US

We present a first–ever comparison of phonetic properties across vocalizations by great apes. We show that “hoot–like” calls by (males of) all non–human great ape genera – chimpanzees, gorillas, and orangutans – overlap with those of human back rounded vowels. Our work underlines the importance of studying the production of calls. Observations from both comparative vocal morphology (non–human great apes have short–and–narrow pharynges and tongues contained in the oral cavity) and observations of vocalizing animals indicate they likely achieve these qualities with disparate articulatory gestures.

## 1. Introduction

All nonhuman great apes produce hoot–like calls, but intra–species comparisons of call properties are rare in the literature. Here, we present the first three–way comparison of properties of hoot–like calls produced by all extant nonhuman great ape genera: chimpanzees (*Pan .spp*), gorillas (*Gorilla .spp*), and orangutans (*Pongo .spp*). We note that across species, hoot–like calls exhibit comparable and overlapping properties; we further suggest that this apparent uniformity results from a derived feature in great apes, i.e., the employment of protruding and rounded lips in call production. We apply the terminology of phonetics and refer to apparent spectral peaks as formants.

## 2. Methods

### 2.1. Formant estimation

To estimate formants, we applied the PREQUEL protocol (Ekström, Moran, Sundberg, & Lameira, 2023). Fundamental frequency ( $f_0$ ) was assessed visually by hand using correlograms (Granqvist & Hammarberg, 2003), and apparent first formant–second formant ( $F_1$ ,  $F_2$ ) coordinates were synthesized and matched for  $f_0$  and compared to the original recording.

### 2.2. Samples

#### 2.2.1. Chimpanzees

Pant hoot calls by Western chimpanzees (*Pan troglodytes verus*) ( $N=50$ , 5 individuals) were recorded by TB at the Taï Chimpanzee Project, Ivory Coast. Pant hoots are divided into four phases, with breathy early-bout lower-frequency vocalizations gradually transitioning into open-mouth high- $f_0$  screams (climaxes) (Grawunder et al., 2022). Because higher-frequency calls are generally nonconductive to formant analysis (Ekström, 2023), only introduction and build-up phases were examined. Pant hoots are performed on both inhalation and exhalation (inbreath, outbreath); we limited analysis to utterances on exhalations (Eklund, 2008). For this study, all sampled individuals were males. This was to control for possible effects of sexual dimorphism – although in comparison with gorillas and orangutans, chimpanzees exhibit relatively little dimorphism (Dixson, 1998).

#### 2.2.2. Gorillas

Western gorilla (*G. gorilla*) silverbacks hoots ( $N=34$ , 2 individuals). Data were recorded by LN at the Bai Hokou and Mongambe field sites in the Dzanga-Sangha Protected Areas in the Central African Republic. Hoots analyzed here are from two adult silverbacks each leading an independent group.

#### 2.2.3. Orangutans

Bornean flanged male orangutan (*P. pygmaeus wurmbii*) long calls (Lameira & Wich, 2008) were sampled and analyzed ( $N=109$ , 9 individuals). Calls were collected at the Tuanan Orangutan Research Station, Central Kalimantan, Borneo, Indonesia by ARL. In our sample data, because recording quality was variable with higher frequencies being lost to high-frequency noise (e.g., birdsong), it was often necessary to segment calls where select portions showed clear and consistent formant frequencies (Ekström et al., 2023).

## 3. Results

Results of our investigations (Figure 1, Table 1) show there is substantial overlap between phonetic properties of great ape hoot-like calls (chimpanzee pant hoot

“hoo’s”, gorilla soft hoots, orangutan long calls) between all three species, and with human close back rounded vowel [u].

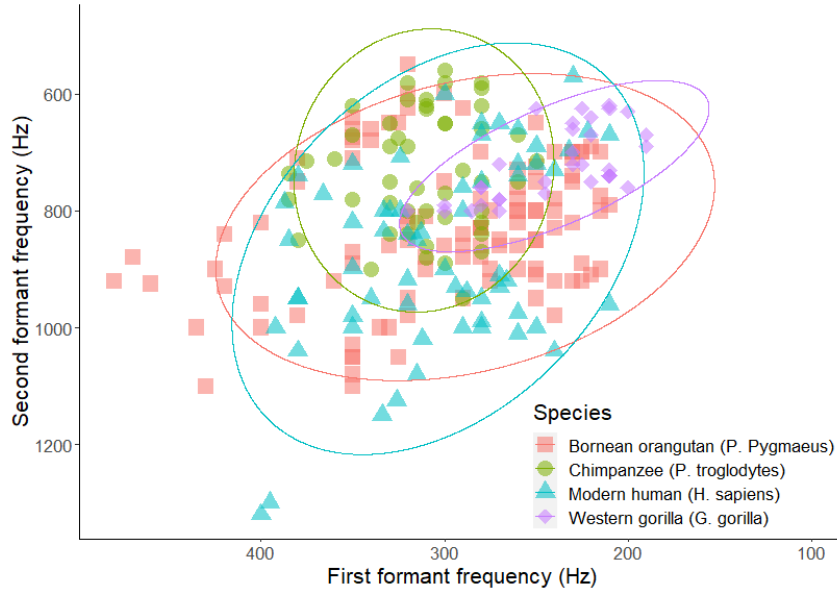


Figure 1. Vowel-like spaces of chimpanzees, gorillas, orangutans producing hoots, and modern human males speaking [u] (Peterson & Barney, 1952). All calls produced by adult males.

Table 1. Mean estimated formants for chimpanzee and gorilla hoots, orangutan long calls, and male human [u]. Values in Hertz. Standard deviations in parentheses.

	Chimpanzee hoots	Gorilla hoots and soft hoots	Orangutan long calls	Human [u]
F <sub>1</sub>	334 (73)	240 (35)	299 (63)	305 (50)
F <sub>2</sub>	748 (108)	721 (60)	829 (120)	871 (159)

We ran linear mixed model analyses in R (ver. 4.3.1) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). We included response as the dependent variable and added random effects of individual animal. Significance was calculated using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017), which applies Satterthwaite’s method to estimate degrees of freedom and generate *P*-values for mixed models. The model specifications were:  $F_n \sim \text{Species} + (1 \mid \text{Subject})$ . Data were log10 transformed prior to analysis. For both F<sub>1</sub> and F<sub>2</sub>, analyses yielded highly significant intercepts ( $P < .001$ ). For F<sub>2</sub>, the effect of the

chimpanzee data was statistically significant ( $P=.012$ ); there were no other significant factor settings. The analysis findings are summarized in Table 2 and Table 3.

Table 2. Linear mixed model for  $F_1$ . Reference level is human [u].

	Estimate	Standard error	df	<i>t</i>	<i>P</i>
(Intercept)	306.12	8.67	55.79	35.32	.00***
Chimpanzee	9	21.46	35.75	.42	.68
Gorilla	-61.41	30.96	29.82	-1.98	.06
Orangutan	15.70	16.76	35.68	.94	.36

Table 3. Linear mixed model for  $F_2$ . Reference level is human [u].

	Estimate	Standard error	df	<i>t</i>	<i>P</i>
(Intercept)	875.91	24.45	47.42	35.82	.00***
Chimpanzee	-167.84	63.84	37.96	-2.62	.01*
Gorilla	-138.91	94.43	34.41	-1.47	.15
Orangutan	2.66	49.91	37.71	.05	.96

## 4. Discussion

### 4.1. *Disparate vocal tract lengths*

Our results suggest that there is substantial overlap between phonetic properties of hoot-like calls across great apes, while species' calls may still be categorically distinguishable from each other. To verify commonalities, we would optimally seek to scale our estimated formants according to vocal tracts length for each species. However, for orangutans and gorillas, no reliable estimates exist. We may, however, take note of some reported findings. Goldstein (1980) measured the length of pharyngeal cavity in an adult human male at 8.9 cm, and length of the oral cavity at 8.1 cm, for a total vocal tract length (VTL) of approximately 17 cm. Nishimura (2005) estimated a total vocal tract length for an adult male chimpanzee at 18.12 cm (computed by adding lengths reported for vertical and horizontal portions of the tract): in terms of total length, the longer chimpanzee face and oral tract compensates for the short pharyngeal tract. Vocal tract lengths of adult humans and chimpanzees thus largely overlap. Further, adult male chimpanzees and orangutans overlap in body size (Dixson, 1998). Because vocal tract length is intimately correlated with body size across primates (Fitch, 1997) – including great apes (Nishimura, 2005) – we may tentatively assume that vocal tract lengths of chimpanzee and orangutan males are likely to overlap also. For these species, biases resulting from differences in vocal tract length are likely to be relatively minor. Notably, however, gorillas represent a significant exception to this

trend, with upper height boundaries of an upright silverback measuring some 30 cm above that of an adult male chimpanzee (Dixson, 1998). Thus, it is likely that while vocal tract lengths for adult male chimpanzees and orangutans may overlap, this is markedly less likely so for silverbacks. Provisioning of great ape vocal tract length data – in particular for gorillas – would significantly improve upon the possibilities to draw conclusions from our findings.

#### 4.2. *Air sacs*

Another imposition to vocal tract length scaling are laryngeal air sacs, the acoustic consequences of which are contested in the literature. de Boer (2012) has suggested air sacs shift up resonances under 2kHz, and introduce an additional low-frequency resonance. We argue that further acoustic modeling efforts, particularly those aimed at exploring the interaction between air sacs and protruding rounded lips, may help resolve this incongruity. Visual inspection of vocalizing animals would also facilitate the modeling of these articulatory behaviors. If air sacs indeed introduce an additional low-frequency resonance, we should treat the apparent  $F_2$  as a “shifted-up”  $F_1$ , and  $F_1$  as a novel resonance induced by the presence of the sacs. This may be consistent with our data. Namely, assuming a VTL = 18.12 cm (Nishimura, 2005), predicted  $F_1$ – $F_2$  of schwa are  $F_1 = 487$  Hz,  $F_2 = 1461$  Hz, according to :

$$Fn = (2n - 1) \cdot c/4 \cdot L \quad (1)$$

where  $n$  is the  $n_{th}$  formant,  $c$  is the speed of sound, and  $L$  is the total length of the tract. Assuming that the articulatory gestures observed by Parr et al. (2005) and Grawunder et al. (2022) for hoots are accurate, we would assume a *longer* tract, as the lips are protruded, effectively shifting down all formants. For example, at VTL = 22 cm, we would expect  $F_1 = 401$  Hz, and  $F_2 = 1203$  Hz, assuming a uniform tube. This is, however, definitively inconsistent with our observations, which put  $F_2$  some  $\approx 450$  Hz below this estimate. Consistent with studies of human vowel production, many studies – including recent efforts by Grawunder et al. (2022) and Ekström et al. (2023) – have focused on measuring and reporting  $F_1$ – $F_2$  dispersion. However, the categorization scheme reported in Grawunder et al. (2022) indicate that, as in human speakers, the apparent first spectral peak is tied to jaw height. Moving forward, we argue that it is necessary to report at least the first three apparent formants, so as to definitively either support or refute the purported roles of air sacs.

#### 4.3. *Articulation*

From an evolutionary perspective, our findings are intriguing, as related vocal tract anatomy and morphology differs significantly between human and nonhuman primates, with nonhuman primates possessing narrow oro– and laryngopharynges

(Negus, 1949, p. 196); the homologous structure in humans is elongated, and open in [u]. There are also likely limitations on intraoral gestures resulting from tongue morphology (Takemoto, 2008). Thus, nonhuman great ape production of [u]-like calls likely involves disparate articulatory gestures (de Boer & Fitch, 2010), likely affecting speech potential more broadly (Ekström & Edlund, 2023a). In addition, speech acoustics modeling exercises indicate that a “two-tube” vocal tract (with proportionate pharyngeal and oral tracts) is more efficient than the standard primate vocal tract (Carré, Lindblom, & MacNeilage, 1995). Observations of apes producing hoots indicate that such calls are often (though not always) produced with comparatively extreme contortions of the lips (Parr, Cohen, & Waal, 2005). Elongating an acoustic chamber will always shift down formants (Fant, 1960), and indeed in chimpanzees, such gestures have been shown to be associated with a reliable shifting down of formants (Grawunder et al., 2022). Our data suggest that the articulatory gestures employed in hoot production – that is, the tendency to affect formant dispersions via the elongation and/or narrowing of the lip passage through rounding – may be a derived feature in nonhuman primates. Understanding morphological aspects involved in the production of apparently similar vowel qualities and vowel-like qualities may yield important insights into critical developments ultimately facilitating the evolution of speech in human ancestors (Ekström & Edlund, 2023b).

### Acknowledgements

The results of this work and the tools used will be made more widely accessible through the national infrastructure Språkbanken Tal under funding from the Swedish Research Council (2017–00626). SM and AE were funded through the Swiss National Science Foundation (PCEFP1\_186841) and AL by the UK Research & Innovation, Future Leaders Fellowship (grant agreement number MR/T04229X/1; ARL). We also thank the Ministère de l’Enseignement supérieur et de la Recherche scientifique; des Eaux et Forêts en Côte d’Ivoire and the Office Ivoirien des Parcs et Réserves, and the Ministry of Higher Education and Scientific Research of Central African Republic (CAR) for permission to conduct this research, the Dzanga–Sangha Protected Areas and WWF CAR for allowing us to carry out fieldwork at their sites, and the UMR 7206 of the National Museum of Natural History in Paris.

### References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- 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.

- de Boer, B. (2012). Loss of air sacs improved hominin speech abilities. *Journal of Human Evolution*, 62, 1–6.
- de Boer, B., & Fitch, T. W. (2010). Computer models of vocal tract evolution: An overview and critique. *Adaptive Behavior*, 18, 36–47.
- Dixon, A. F. (1998). *Primate sexuality: Comparative studies of the prosimians, monkeys, apes, and human beings*. Oxford, UK: Oxford University Press.
- Eklund, R. (2008). Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association*, 38, 235–324.
- Ekström, A. G. (2023). Ape vowel-like sounds remain elusive: A comment on Grawunder et al. (2022). *International Journal of Primatology*, 44, 237–239.
- Ekström, A. G., & Edlund, J. (2023a). Evolution of the human tongue and emergence of speech biomechanics. *Frontiers in Psychology*, 14, 2132.
- Ekström, A. G., & Edlund, J. (2023b). Sketches of chimpanzee (*Pan troglodytes*) hoo's: vowels by any other name? *Primates*, 65, 81–88.
- 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* (pp. 3076–3080). Prague, Czech Republic: GUARANT International spol. s r.o.
- Fant, G. (1960). *The acoustic theory of speech production*. The Hague: Mouton.
- Fitch, T. W. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, 102, 1213–1222.
- Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children*. [Doctoral Dissertation]. Massachusetts Institute of Technology.
- Granqvist, S., & Hammarberg, B. (2003). The correlogram: A visual display of periodicity. *The Journal of the Acoustical Society of America*, 114, 2934–2945.
- Grawunder, S., Uomini, N., Samuni, L., Bortolato, T., Girard-Buttoz, C., R. M. 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, 20200455.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
- 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.

- Negus, V. E. (1949). *Comparative anatomy and physiology of the larynx*. Heine-mann.
- Nishimura, T. (2005). Developmental changes in the shape of the supralaryngeal vocal tract in chimpanzees. *American Journal of Physical Anthropology*, 126, 193–204.
- Parr, L. A., Cohen, M., & Waal, F. D. (2005). Influence of social context on the use of blended and graded facial displays in chimpanzees. *International Journal of Primatology*, 26, 73–103.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24, 175–184.
- Takemoto, H. (2008). Morphological analyses and 3d modeling of the tongue musculature of the chimpanzee (pan troglodytes). *American Journal of Primatology*, 70, 966–975.