



OPEN

## Open plains are not a level playing field for hominid consonant-like versus vowel-like calls

Charlotte Gannon<sup>1</sup>✉, Russell A. Hill<sup>2,3,4</sup> & Adriano R. Lameira<sup>1</sup>✉

Africa's paleo-climate change represents an "ecological black-box" along the evolutionary timeline of spoken language; a vocal hominid went in and, millions of years later, out came a verbal human. It is unknown whether or how a shift from forested, dense habitats towards drier, open ones affected hominid vocal communication, potentially setting stage for speech evolution. To recreate how arboreal proto-vowels and proto-consonants would have interacted with a new ecology at ground level, we assessed how a series of orangutan voiceless consonant-like and voiced vowel-like calls travelled across the savannah. Vowel-like calls performed poorly in comparison to their counterparts. Only consonant-like calls afforded effective perceptibility beyond 100 m distance without requiring repetition, as is characteristic of loud calling behaviour in nonhuman primates, typically composed by vowel-like calls. Results show that proto-consonants in human ancestors may have enhanced reliability of distance vocal communication across a canopy-to-ground ecotone. The ecological settings and soundscapes experienced by human ancestors may have had a more profound impact on the emergence and shape of spoken language than previously recognized.

Over the last 17 million years, continental tectonic movement<sup>1</sup> combined with global cooling and aridification resulted in the spread of grasslands across Eurasia and Africa<sup>2</sup>, ultimately setting the stage for the emergence of the human lineage<sup>3,4</sup>. Throughout these continents' landmass, dense mixed forests<sup>5</sup> gave way to generally open and drier habitats as new regional but complex patterns of wet-dry cycles became established<sup>6-8</sup>. These extreme climate changes imposed new ecological conditions for hominid survival, which resulted in major great ape extinctions starting at 9.5 million years ago in Eurasia and Africa<sup>9-12</sup>. While the transformations imposed on hominid anatomy and behaviour can be directly reconstructed from archaeological sites<sup>4,13</sup> or inferred from paleo-ecology<sup>9,14,15</sup>, one major component of human evolution is unrecoverable from the fossil record—the evolution of spoken language and its precursor, vocal signals.

Paleontological evidence can help understand the anatomical changes that facilitated the expression of language through speech, such as larynx's descent<sup>16</sup>, air-sacs' recession<sup>17-19</sup>, thoracic innervations' increase<sup>20,21</sup> and mandibular transformations<sup>22,23</sup>. However, these structures are remote surrogates of the signals used to de facto communicate vocally, and ultimately, to articulate (proto)language. These structures have also been tempered by major selective forces driving other changes in human anatomy, such as head and vocal tract re-organization due to a rising bipedal posture<sup>24</sup> and diet changes<sup>23</sup>, overall, making it difficult to parse language-causes versus language-consequences based on anatomy or the fossil record alone.

Being our closest living relatives, nonhuman great apes represent extant models of hominids' ancestral biology and behaviour<sup>10</sup>, including homologous sound production<sup>25</sup> and sound perception capacities<sup>26</sup>. They provide an alternative and potent means to explore why, among its contemporary and sympatric species, only the repertoire of ancient hominids took a turn towards spoken language in the wake of continent-wide ecological transformations, from densely forested habitats to drier and more open soundscapes.

### Great ape proxies of proto-vowels and proto-consonants

All great ape genera produce voiceless consonant-like calls<sup>25,27-35</sup> and voiced vowel-like calls<sup>36-39</sup>, in direct articulatory and acoustic homology with all the world's languages. *Pan* (chimpanzees and bonobos) and *Pongo* (orangutans), can produce, respectively, one and both of these call categories at speech-like rhythm<sup>32,40</sup>. Given that each and every spoken language is composed by consonants and vowels that combine to compose a language's

<sup>1</sup>Department of Psychology, University of Warwick, Coventry, UK. <sup>2</sup>Department of Anthropology, Durham University, Durham, UK. <sup>3</sup>Primate and Predator Project, Soutpansberg Mountains, Thohoyandou, South Africa. <sup>4</sup>Department of Biological Sciences, University of Venda, Thohoyandou, South Africa. ✉email: charlotte.gannon@warwick.ac.uk; adriano.lameira@warwick.ac.uk

words and sentences, great ape consonant-like and vowel-like calls offer desirable proxies of the two putative speech precursor elements.

Among great apes, orangutans exhibit a particularly large consonant-like repertoire<sup>41</sup> that they combine with various vowel-like voiced calls to produce syllable-like combinations<sup>36</sup> in complex sequences<sup>28,42,43</sup>. Consonant-like and vowel-like calls exhibit unique acoustic profiles that result from distinct mechanics and control, and in orangutans, both have been shown to be under fine motor laryngeal<sup>44–46</sup> and supra-laryngeal control<sup>30,31</sup>. Similarly to the speech sounds, both call categories can occur as population-specific traditions in orangutans instead of representing calls that are universal in the species, and their articulation and acoustics are also socially moulded<sup>47–52</sup>. In addition, wild orangutans use these combinations as a canvas for the expression of high cognitive processes, such as tool-modulated vocal deception<sup>28</sup> and vocal communication about past events<sup>53</sup>.

To address the role of paleo-climate change on spoken language evolution, African apes, for instance, could allow comparison of the same vowel-like call between forest and savannah populations of the same species or genus. However, orangutans are the only great ape currently known to produce syllable-like call combinations in the wild, making these structures the only extant models that can be considered as combinatorial homologues of proto-linguistic structures in human ancestors. Furthermore, orangutans are the most arboreal of all great apes<sup>54</sup>, providing a rare opportunity to explore the “canopy-to-ground” ecological shift experienced by ancient hominids and human ancestors. Indeed, cumulating evidence shows that human ancestors were more arboreal than traditionally appreciated, and more than extant African apes<sup>41,55–58</sup>.

### Re-staging the ecology of spoken language evolution

By taking advantage of orangutans’ arboreal consonant-like and vowel-like calls and moving them to an open landscape setting, we can recreate, as close and realistically as it is possible today, the scenario in the Middle and Late Miocene (16–5.3 mya) when hominids transitioned down from trees to an open landscape. This is not to suggest that modern day consonants, vowels, or syllables directly descended from orangutans<sup>58,59</sup>. Instead, orangutans’ speech-like calls and combinations offer extant homologues to study extinct proto-linguistic units and structures.

For spoken language to evolve, consonant-like and vowel-like calls had to marry to give birth to proto-syllables, -words and -sentences. While vowel-like calls are prototypical to all nonhuman primates and well represented in the call repertoires of all great apes, consonant-like calls are predominantly found in arboreal versus ground-dwelling great apes<sup>41</sup>. This implies that, whenever and wherever human ancestors descended from trees during the Mid and Late Miocene, they brought with them an arboreal repertoire, predicted to have been rich in consonant-like calls, in addition to primates’ prototypical vowel-like repertoire. Understanding what this transition meant for the use of the two foundational elements for speech is, hence, critical for an authentic reconstruction of the speech’s evolutionary process and origin of language-able apes.

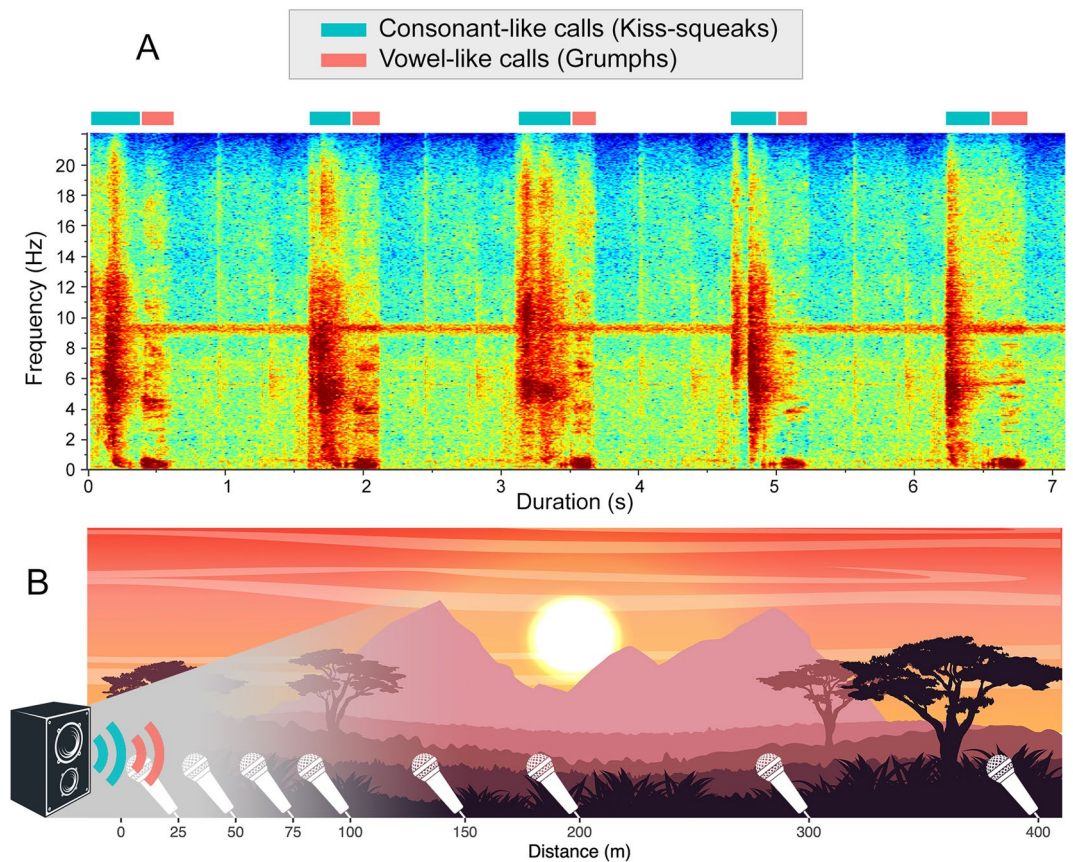
To shed new light on the ecology of spoken language evolution and the most dramatic climate transition known to have occurred alongside the divergence of human ancestors from other hominid lineages, we conducted playbacks of orangutan voiceless consonant-like ‘kiss-squeaks’<sup>36</sup> produced in combination with voiced vowel-like ‘grumphs’<sup>36</sup> across an African savannah. We then quantified their respective acoustic performance by recording played back calls at increasing distances up to 400 m (Fig. 1, Supplementary data 1). Played back calls were originally recorded across various individuals, various populations and under three different contexts (towards observers, towards a white predator-model and towards a tiger-patterned predator-model)<sup>43,49,60,61</sup>. Previous research has established that orangutan consonant-like and vowel-like calls can both be information-dense<sup>62</sup> and that they both broadcast and transmit information successfully across a dense rainforest up to 100 m distance<sup>63</sup>. However, how the two call categories behave and interact in open landscapes with fewer to no physical obstructions against signal transmission, such as the savannah, remains unknown. The classic inverse square law of sound propagation for pure tones predicts that low frequencies degrade less over distance than higher frequencies. Accordingly, vowel-like calls, which typically show lower frequency contours than consonant-like calls, would be theoretically predicted to carry further and better. Nonetheless, previous bioacoustics research has shown that theoretical models are poor substitutes for real, living biological calls with acoustic profiles more complex than pure tones<sup>63</sup>. The question of whether consonant-like and vowel-like calls behaved equally in the emergent ecology of open landscapes during human evolution remains, thus, open.

### Results

The acoustic performance of both voiceless consonant-like kiss-squeaks (N = 487 at 0 m distance) and voiced vowel-like grumphs (N = 487 at 0 m distance) changed significantly over the 400 m of savannah (Fig. 2, Supplementary data 2). That is, delta time, max frequency, max slope, signal-to-noise ratio (SNR) and max amplitude all significantly degraded (or presented absence thereof) with increasing distance for both kiss squeaks (LMM ANOVA,  $p < 0.001$  for all parameters; Supplementary data 2) and grumphs (LMM ANOVA,  $p < 0.001$  for all parameters; Supplementary data 3).

In addition to an effect over distance, population and/or context had a significant effect on several acoustic parameters for kiss squeaks (LMM ANOVA, delta time/population:  $p < 0.02$ ; delta time/context:  $p < 0.001$ ; max frequency and max slope/context:  $p < 0.001$ ; SNR/population:  $p = 0.003$ ; SNR/context:  $p = 0.0037$ ; max amplitude/context:  $p < 0.001$ ; Supplementary data 2). Context also had a significant effect on several acoustic parameters for grumphs (LMM ANOVA, delta time and max slope:  $p < 0.001$ ; max amplitude:  $p = 0.003$ ; Supplementary data 3).

In terms of overall detection or “acoustic survival”—that is, whether calls were audibly perceptible and visible on the spectrogram or not—vowel-like calls’ survival was significantly compromised and significantly lower than consonant-like calls from 100 m onwards. That is, there was no overlap between bands of 95% confidence interval of both calls after this distance (Fig. 3). After a sharp decay around 125 m away, less than 20% of vowel-like call



**Figure 1.** Spectrographic representation of orangutan consonant-like and vowel-like calls (above) and experimental set up (below).

survived up to 400 m distance. Conversely, consonant-like calls exhibited only a modest decrease in survival from 250 m onwards, with ~80% of the calls surviving (i.e., remaining perceptible) up to 400 m distance.

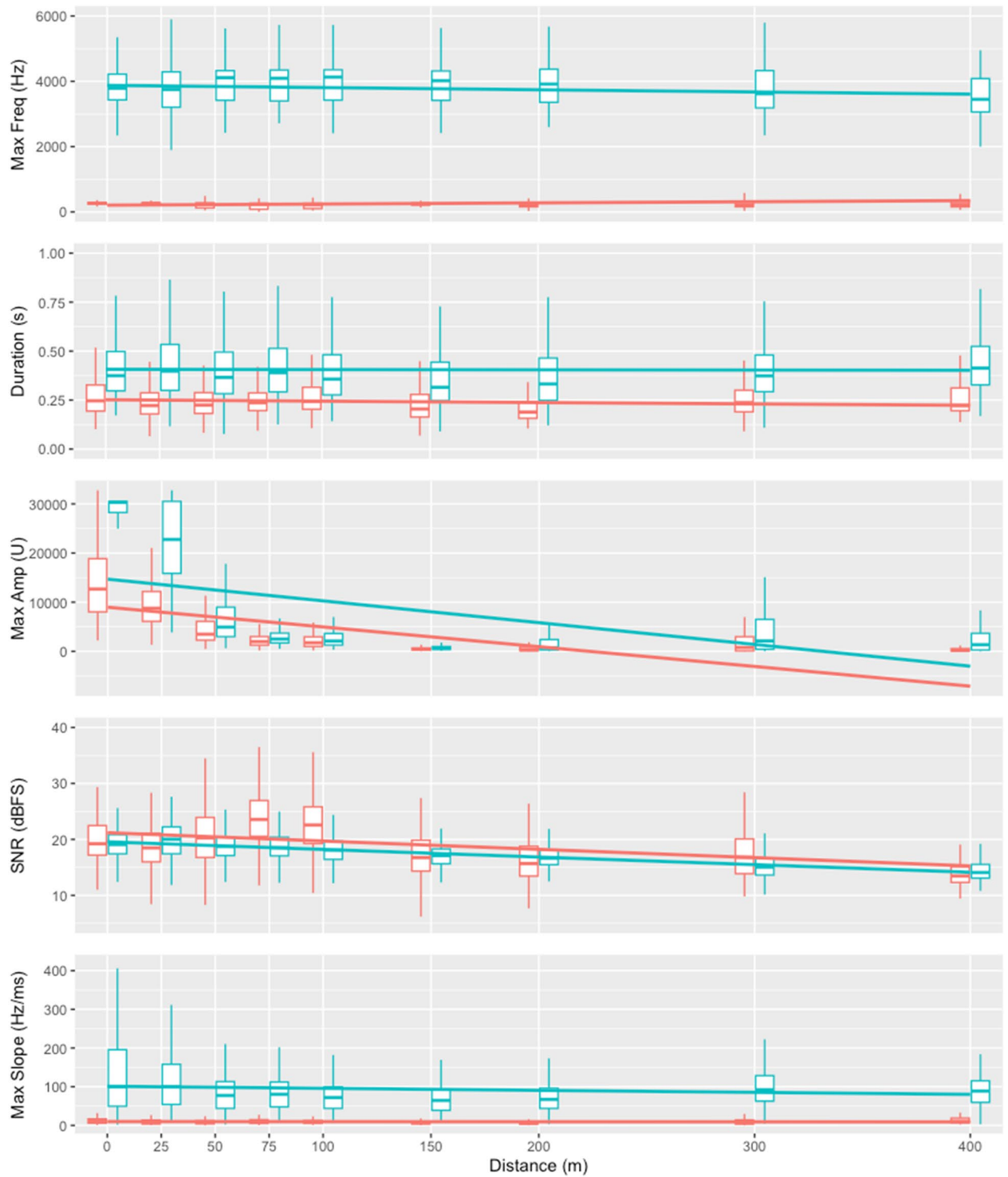
## Discussion

We established that consonant-like and vowel-like calls produced in combination by a wild arboreal great ape degrade differently across a savannah open plain. Notably, orangutan vowel-like grumphs exhibited a significantly lower acoustic performance than orangutan consonant-like kiss-squeaks, with consonant-like calls significantly outperforming their vowel-like counterparts beyond distances of 100 m. At 400 m vowel-like calls were largely inaudible (<20% survival rate), whereas most consonant-like calls were still largely detectable and perceptible (>80% survival rate). Additional population and context effects were not unexpected based on known geographic and contextual variation encoded in orangutan vocal signals<sup>62</sup> and primate call communication more generally.

Our setup involved a *series* of calls played back every few seconds for roughly 15 min. While this was ideal to assess sound propagation over distance, as classically used in bioacoustics<sup>61,63–66</sup>, the repetition of calls probably inflated the chances of call detection during the phase of extraction of acoustic parameters. This may have particularly inflated the detection of faint or partly inaudible vowel-like calls, given they had a rapid acoustic degradation and poor survival over distance (i.e., acoustic survival = 0% at 200, 300 and 400 m; Fig. 3; Supplementary data 3). Conversely, repetition of consonant-like calls seems to have been largely redundant given that their acoustic survival rates were largely immune to distance. If we had played back solo or short call series (as kiss-squeaks and grumphs frequently occur naturally), it could have been possible that vowel-like calls would be fully imperceptible beyond 125 m.

## Implications for signal and information theory

Results show that the law of sound propagation for pure tones, which predicts that lower-frequency sounds (characteristic of voiced vowel-like calls) travel farther than higher-frequency (characteristic of voiceless consonant-like calls)<sup>67</sup>, does not apply to natural calls. Natural calls are not pure tones, but instead, composed by a varied mix of wide sound bands, particularly voiceless calls. These are typically noisy and present, hence, a wide spread of acoustic energy across the frequency spectrum. Natural calls with complex and composite acoustic profiles behave differently as a whole than defined, “clean” frequencies. This made our results the opposite of what would be theoretically expected for pure tones; the acoustic survival of higher-frequency consonant-like calls was higher than lower-frequency vowel-like calls. This speaks to the importance of using great ape calls as extant models

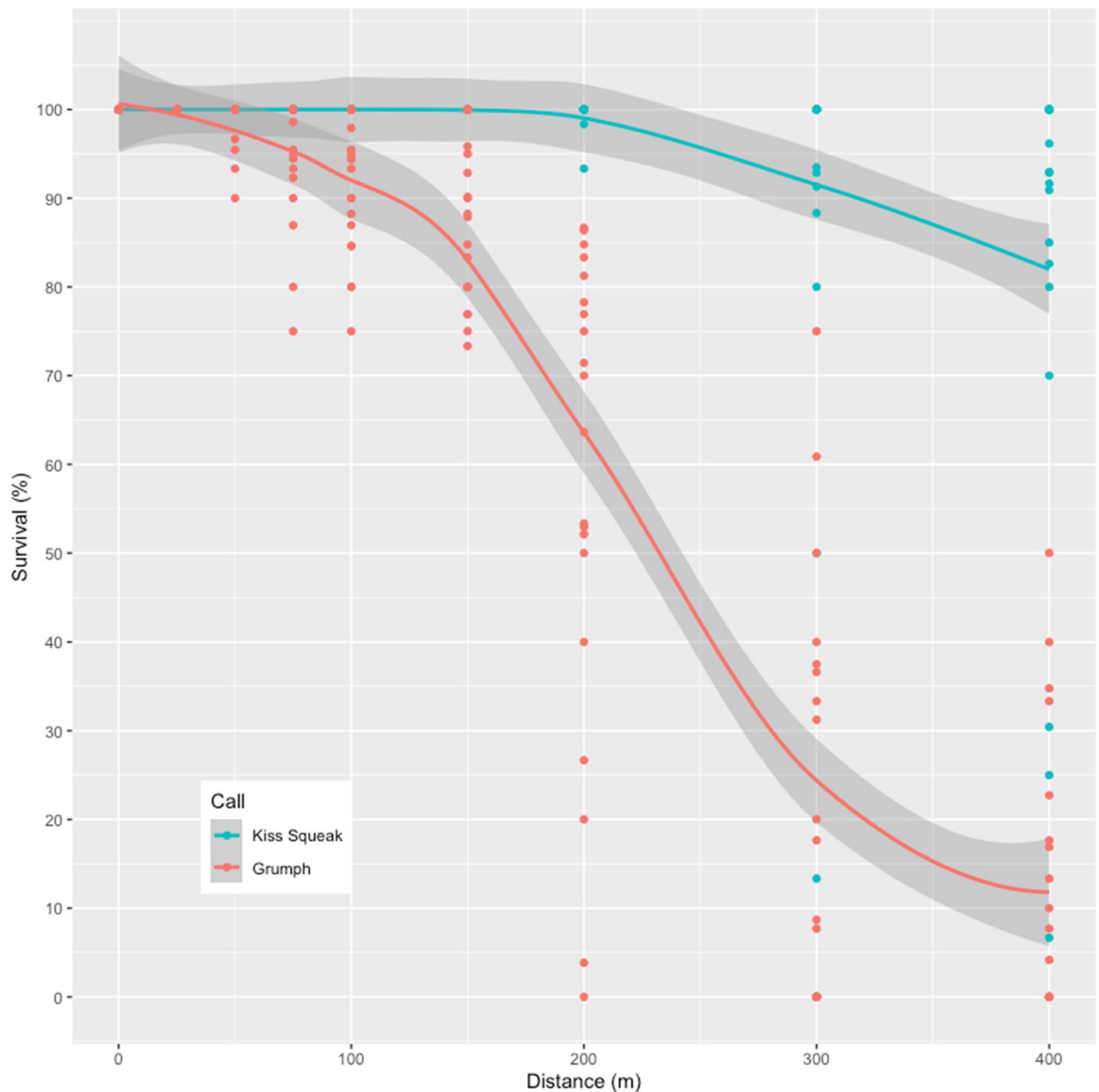


**Figure 2.** Acoustic performance over distance. Box plots represent median and 25–75% interquartile range, whiskers represent lower/highest value within 1.5 times interquartile range below/above, outliers omitted strictly for visual clarity (they were retained in all analytical procedures). Linear trend lines represented across distance are for visual aid only (based on raw data, not model projections).

for spoken language origin and evolution, despite the common use of purely theoretical, computational, and/or mathematical models in the field<sup>68</sup>.

### Evolutionary implications

Results show that paleo-climate change during the Middle and Late Miocene put forth new selection forces for hominid consonant-like versus vowel-like calls. Where dense forests had offered equivalent performance for



**Figure 3.** Acoustic survival over distance. Perceptibility of consonant-like and vowel-like calls across increasing savannah distances up to 400 m, 95% confidence interval (shaded area).

both call categories<sup>63</sup>, the newly emerging dry and open landscapes offered superior transmission efficiency to consonant-like calls over mid and long distances. Until the time of these ecological transformations, primates in general, and hominids in particular, had relied for millions of years on the production of repetitive, strenuous, and vigorous loud vocal displays for distance communication, composed by voiced vowel-like calls<sup>69</sup>. In the new ecological settings, hominids now had an element available—consonant-like calls—in their vocal repertoire with enhanced perceptibility, if and when required.

Enhanced perceptibility of consonants versus vowels at the earliest stages of speech evolution could help to start explaining some features of modern spoken languages. For instance, consonants can act as natural cues in speech, preferably used by language learners for perceptually breaking sentences and inserting processing pauses where there are none<sup>70</sup>. Also, during the babbling stage of an infant's language acquisition and development, the larger the range of consonants they are exposed to, the earlier they will typically begin to babble and reach speech development milestones<sup>71</sup>. Infants also learn to rely on consonants more than vowels to identify words after their first year<sup>72</sup>. Consonants and vowels then continue to play different roles for language users as adults; consonants are relied on to extract cues primarily related to semantic information, whereas vowels are relied on to extract cues primarily related to syntax<sup>73</sup>. All these linguistic phenomena depend on heightened conspicuity of consonant sounds. Accordingly, it is possible that consonants' growing role in hominid vocal communication

started to play out in the new land- and soundscapes that human ancestors encountered in the wake of paleo-climate change. The ecology of ancient hominids may have moulded human modern verbal behaviour to a larger extent than hitherto appreciated.

## Methods

### In brief

Consonant-like and vowel-like calls that had been previously recorded in the form of syllable-like combinations from wild orangutans were played back across a savannah habitat in South Africa and re-recorded over increasing distance. Five acoustic parameters were extracted from the recordings at distance intervals from zero to 400 m away. The degradation of these parameters was used to evaluate the acoustic performance of consonant-like and vowel-like calls. Individual, contextual, and geographical variation among consonant-like and vowel-like calls, and the capacity of a receiver to assign a call to its correct class, was used to evaluate information broadcast over distance.

### Study site

Playback experiments were conducted at Lajuma in the western Soutpansberg Mountain Range, Limpopo Province, South Africa (23°06'45.14" S, 29°11'37.10" E). The playback took place in a grassland with woodland patches at elevation of 1420m<sup>74</sup>. The wider Soutpansberg's geography presents a complex environment, from mist-belt forest groups and closed woodlands, to bushveld complex and rocky mountainous regions<sup>75</sup>.

### Data collection

The playback call set was made up of 487 calls from populations of Sumatran (*Pongo abelii*) and Bornean orangutans (*Pongo pygmaeus*); playbacks included calls from 20 different individuals, three different contexts, and three different geographical populations<sup>63</sup>. The recordings were collected from three research stations: Gunung Palung and Tuanan in West and Central Kalimantan, Indonesian Borneo, and in Sampan Getek in North Sumatra, Indonesia. All kiss-squeaks and grumphs selected for playback were produced in kiss-squeak + grumph call combinations. In preparation for playback, all call combinations were set to the same peak amplitude using Raven interactive sound analysis software (v. 1.6, Cornell Lab of Ornithology, Ithaca, New York). In other words, we ensured that the loudest point of each call combination was as loud as the others. This preserved the natural difference in loudness between the call types within the same combination.

The playback speaker was set at 0 m and calls were re-recorded every 25 m up to 400 m along a transect across flat and dry terrain. All playbacks were conducted on the same morning and meteorological conditions remained consistent during the experiment. There had been no heavy wind or rainfall in the previous 48 h of recording. Theoretically, wind and rain were assumed to be adverse to sound transmission across habitats, including forests for example<sup>63</sup>, not just open habitats. Playbacks were conducted once per distance using a Marantz Digital Recorder PMD-660 (D&M Holdings, Kawasaki, Japan) connected to a Nagra DSM speaker (Audio Technology Switzerland S.A., Romanel, Switzerland). The base of the speaker was set between 1 and 1.5 m from the ground. The playbacks were re-recorded using a ZOOM H4next Handy Recorder (ZOOM Corporation, Tokyo, Japan) connected to a RØDE NTG-2 directional microphone (RØDE LLC, Sydney, Australia). Audio data was recorded using the WAVE PCM format at 16 bits. The mic was set parallel to the ground from 1 to 1.5 m. All volume settings for the mic and speakers remained equal for all distances.

### Data measurements

We extracted acoustic parameters per call across each distance of 25, 50, 75, 100, 150, 200, 300 and 400 m away from the speaker. Data from 0 m was extracted from the original playback call set. Acoustic parameters per call were extracted by manually drawing a selection box in Raven around a call or whichever call components were still visible. If no trace of a call was visible in the spectrogram, no selection was drawn and no parameters were extracted, given that a call was effectively inaudible in these cases. A call was then considered to not have survived at that distance. All selections were drawn by the same experimenter to avoid inter-observer biases. Five acoustic parameters were measured from all calls using Raven interactive sound analysis: delta time (s), max frequency (Hz), peak frequency counter slope (Hz/ms), signal-to-noise ratio (SNR) (dB FS) and max amplitude (u) (Supplementary data pages 1–16). Delta time is the time difference between the offset and onset of each call. Max frequency is the frequency with the highest amplitude in a call. PFC max slope is the increase or decrease of the slope for each call. SNR is the signal power from the call compared to background. Max amplitude is the maximum power of the sound wave for each call. These parameters were used as previous research has demonstrated that they are able to provide strong descriptors of acoustic and informational context<sup>49</sup>. Each of the 5 parameters could be extracted from both the kiss squeaks and grumphs, enabling direct comparison between the acoustic and informational performance across distances for both call types.

### Data analyses

To assess acoustic performance, Linear Mixed Models (model type: III sum of squares; test model terms: Satterthwaite) were conducted using JASP (v. 0.15). One model was generated for each acoustic parameter ( $\times 5$ ; i.e., time, max freq, max slope, SNR and max amp) under both call type conditions ( $\times 2$ ; i.e., kiss-squeak and grumph), resulting in 10 models in total. For each model, the acoustic parameter was inserted as dependent variable ( $N = 4278$  kiss-squeaks,  $N = 3302$  grumphs). Distance (ordinal: 0, 25, 50, 75, 100, 150, 200, 300, 400 m), context (three levels: calls produced towards observers, towards a white predator-model, towards a tiger-patterned predator-model) and geographical population (three levels: Tuanan, Gunung Palung, Sampan Getek) were

inserted as fixed effect variables. Context and population were included in analyses of acoustic performance to maintain the same call sample used for information performance analyses.

Additionally, name of individual (N = 20) was inserted as random effect, as some individuals were re-used across the 3 different contexts. Call ID (33 levels) was inserted as random effect, given that the same calls were re-recorded per distance. LMM graphs were created in R using ggplot2 and gridExtra. For grumph's max frequency, we obtained a warning that model fit was singular, and so, for this model, we removed call ID as random effect, which yielded the same qualitative results. To graphically depict the “acoustic survival” of each call type, we used ggplot2 in R to render the percentage of calls heard per distance.

## Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the electronic supplementary materials. Audio files are available on the Open Science Framework ([https://osf.io/6hkaq/?view\\_only=048983a0c3744cb9814431e1b084dc2c](https://osf.io/6hkaq/?view_only=048983a0c3744cb9814431e1b084dc2c)). Additional data may be requested from the corresponding author, CG, upon reasonable request.

Received: 27 July 2023; Accepted: 22 November 2023

Published online: 21 December 2023

## References

1. Wichura, H. *et al.* A 17-My-old whale constrains onset of uplift and climate change in east Africa. *PNAS* **112**, 3910–3915 (2015).
2. Kaya, F. *et al.* The rise and fall of the Old World savannah fauna and the origins of the African savannah biome. *Nat. Ecol. Evol.* **2**, 241–246 (2018).
3. Scerri, E. M. L., Chikhi, L. & Thomas, M. G. Beyond multiregional and simple out-of-Africa models of human evolution. *Nat. Ecol. Evol.* **3**, 1370–1372 (2019).
4. Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E. M. L. & Skoglund, P. Origins of modern human ancestry. *Nature* **590**, 229–237 (2021).
5. Michel, L. A. *et al.* Remnants of an ancient forest provide ecological context for Early Miocene fossil apes. *Nat. Commun.* **5**, 1 (2014).
6. Maslin, M. A. *et al.* East African climate pulses and early human evolution. *Quat. Sci. Rev.* **101**, 1–17 (2014).
7. deMenocal, P. B. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth Planet. Sci. Lett.* **220**, 3–24 (2004).
8. Blinkhorn, J., Timbrell, L., Grove, M. & Scerri, E. M. L. Evaluating refugia in recent human evolution in Africa. *Phil. Trans. R. Soc. B* **377**, 20200485 (2022).
9. Potts, R. Paleoenvironmental basis of cognitive evolution in great apes. *Am. J. Primatol.* **62**, 209–228 (2004).
10. Lameira, A. R. & Call, J. Understanding language evolution: Beyond *pan*-centrism. *BioEssays* **42**, 1900102 (2020).
11. Raia, P. *et al.* Past extinctions of homo species coincided with increased vulnerability to climatic change. *One Earth* **3**, 480–490 (2020).
12. Timmermann, A. *et al.* Climate effects on archaic human habitats and species successions. *Nature* **604**, 495–501 (2022).
13. Scerri, E. M. *et al.* Did Our species evolve in subdivided populations across Africa, and why does it matter?. *Trends Ecol. Evol.* **33**, 582 (2018).
14. Potts, R. *et al.* Increased ecological resource variability during a critical transition in hominin evolution. *Sci. Adv.* **6**, 8975 (2020).
15. Maslin, M. A. & Christensen, B. Tectonics, orbital forcing, global climate change, and human evolution in Africa: Introduction to the African paleoclimate special volume. *J. Hum. Evol.* **53**, 443–464 (2007).
16. Nishimura, T., Mikami, A., Suzuki, J. & Matsuzawa, T. Descent of the larynx in chimpanzee infants. *Proc. Natl. Acad. Sci.* **100**, 6930–6933 (2003).
17. Boer, B. Modelling vocal anatomy's significant effect on speech. *J. Evol. Psychol.* **8**, 351–366 (2010).
18. Boer, B. Loss of air sacs improved hominin speech abilities. *J. Hum. Evol.* **62**, 1–6 (2012).
19. Boer, B. Air sacs and vocal fold vibration: Implications for evolution of speech. *Theor. Hist. Scient.* **9**, 13–28 (2012).
20. Ann, M. & Hewitt, G. Increased breathing control: Another factor in the evolution of human language. *Evol. Anthropol. Issues News Rev.* **13**, 181–197 (2004).
21. MacLarnon, A. M. & Hewitt, G. P. The evolution of human speech: The role of enhanced breathing control. *Am. J. Phys. Anthropol.* **109**, 341–363 (1999).
22. Bermejo-Fenoll, A., Panchón-Ruiz, A. & Walker, M. J. Neanderthal, Sapiens, and chimpanzee mandibles: A comparative study in relation to articulated speech. *Archaeol. Anthropol. Sci.* **14**, 116 (2022).
23. Blasi, D. *et al.* Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science* <https://doi.org/10.1126/science.aav3218> (2019).
24. Lieberman, D. *The Evolution of the Human Head* (Belknap Press of Harvard University Press, 2011).
25. Lameira, A. R., Maddieson, I. & Zuberbühler, K. Primate feedstock for the evolution of consonants. *Trends Cogn. Sci.* **18**, 60–62 (2014).
26. Quam, R. *et al.* Early hominin auditory capacities. *Sci. Adv.* **1**, e1500355 (2015).
27. Lameira, A. R. The forgotten role of consonant-like calls in theories of speech evolution. *Behav. Brain Sci.* **37**, 559–560 (2014).
28. Hardus, M. *et al.* Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation?. *Proc. R. Soc. B.* <https://doi.org/10.1098/rspb.2009.1027> (2009).
29. Lameira, A. R. *et al.* Population-specific use of the same tool-assisted alarm call between two wild orangutan populations (*Pongo pygmaeus wurmbii*) indicates functional arbitrariness. *PLoS One* **8**, e69749 (2013).
30. Lameira, A. R. *et al.* Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *J. Acoust. Soc. Am.* **134**, 1–11 (2013).
31. Wich, S. *et al.* A case of spontaneous acquisition of a human sound by an orangutan. *Primates* **50**, 56–64 (2009).
32. Lameira, A. R. *et al.* Speech-like rhythm in a voiced and voiceless orangutan call. *PLoS One* **10**, e116136 (2015).
33. Watts, D. P. Production of grooming-associated sounds by chimpanzees (*Pan troglodytes*) at Ngogo: Variation, social learning, and possible functions. *Primates* **2015**, 1–12. <https://doi.org/10.1007/s10329-015-0497-8> (2015).
34. Perlman, M. & Clark, N. Learned vocal and breathing behavior in an enculturated gorilla. *Anim. Cogn.* **18**, 1165–1179 (2015).
35. Lameira, A. R. & Moran, S. Life of p: A consonant older than speech. *BioEssays* **2023**, 200246. <https://doi.org/10.1002/bies.20220246> (2023).
36. Hardus, M. E. *et al.* A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation. In *Orangutans* (Oxford University Press, 2008). <https://doi.org/10.1093/acprof:oso/9780199213276.003.0004>.

37. Grawunder, S. *et al.* Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. *Phil. Trans. R. Soc. B* **377**, 20200455 (2022).
38. Salmi, R., Hammerschmidt, K. & Diane, M. D. S. Western Gorilla vocal repertoire and contextual use of vocalizations. *Ethology* **119**, 114 (2013).
39. Grawunder, S. *et al.* Higher fundamental frequency in bonobos is explained by larynx morphology. *Curr. Biol.* **28**, R1188–R1189 (2018).
40. Pereira, A. S., Kavanagh, E., Hobaiter, C., Slocombe, K. E. & Lameira, A. R. Chimpanzee lip-smacks confirm primate continuity for speech-rhythm evolution. *Biol. Lett.* **16**, 20200232 (2020).
41. Lameira, A. R. Arboreal origin of consonants and thus, ultimately, speech. *TiCS* **27**, 122–124 (2022).
42. Kershenbaum, A. *et al.* Animal vocal sequences: Not the Markov chains we thought they were. *Proc. Biol. Sci. R. Soc.* **281**, 20141370 (2014).
43. Lameira, A. R. *et al.* Predator guild does not influence orangutan alarm call rates and combinations. *Behav. Ecol. Sociobiol.* **67**, 519–528 (2013).
44. Lameira, A. R. & Shumaker, R. W. Orangutans show active voicing through a membranophone. *Sci. Rep.* **9**, 12289 (2019).
45. Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A. & Shumaker, R. W. Vocal fold control beyond the species-specific repertoire in an orang-utan. *Sci. Rep.* **6**, 30315 (2016).
46. Lameira, A. R. & Hardus, M. E. Wild orangutans can simultaneously use two independent vocal sound sources similarly to song-birds and human beatboxers. *PNAS Nexus* **2**, 182 (2023).
47. Lameira, A. R. *et al.* Sociality predicts orangutan vocal phenotype. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-022-01689-z> (2022).
48. Wich, S. A. *et al.* Call cultures in orang-utans?. *PLoS One* **7**, e36180 (2012).
49. Lameira, A. R. *et al.* Proto-consonants were information-dense via identical bioacoustic tags to proto-vowels. *Nat. Hum. Behav.* **1**, 1–5 (2017).
50. Watson, S. K. *et al.* Vocal learning in the functionally referential food grunts of chimpanzees. *Curr. Biol.* **25**, 495–499 (2015).
51. Crockford, C., Herbinger, I., Vigilant, L. & Boesch, C. Wild chimpanzees produce group-specific calls: A case for vocal learning?. *Ethology* **110**, 221–243 (2004).
52. Lameira, A. R., Delgado, R. & Wich, S. Review of geographic variation in terrestrial mammalian acoustic signals: Human speech variation in a comparative perspective. *J. Evol. Psychol.* **8**, 309–332 (2010).
53. Lameira, A. R. & Call, J. Time-space-displaced responses in the orangutan vocal system. *Sci. Adv.* **4**, 3401 (2018).
54. Delgado, R. A. & van Schaik, C. P. The behavioral ecology and conservation of the orangutan (*Pongo pygmaeus*): A tale of two islands. *Evol. Anthropol. Issues News Rev.* **9**, 201–218 (2000).
55. Drummond-Clarke, R. C. *et al.* Wild chimpanzee behavior suggests that a savanna-mosaic habitat did not support the emergence of hominin terrestrial bipedalism. *Sci. Adv.* **8**, 9752 (2022).
56. Kappelman, J. *et al.* Perimortem fractures in Lucy suggest mortality from fall out of tall tree. *Nature* <https://doi.org/10.1038/nature19332> (2016).
57. Thorpe, S., Holder, R. & Crompton, R. Origin of human bipedalism as an adaptation for locomotion on flexible branches. *Sci. N. Y. N. Y.* **316**, 1328–1331 (2007).
58. White, T. D., Lovejoy, O. C., Asfaw, B., Carlson, J. P. & Suwa, G. Neither chimpanzee nor human, *Ardipithecus* reveals the surprising ancestry of both. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4877–4884 (2015).
59. Barton, R. Chimp comparisons won't explain human evolution. *Nature* **443**, 632 (2006).
60. Lameira, A. R. & Call, J. Time-space-displaced responses in the orangutan vocal system. *Science Advances* **4**, 11 (2018).
61. Lameira, A. R. & Wich, S. Orangutan long call degradation and individuality over distance: A playback approach. *Int. J. Primatol.* **29**, 615–625 (2008).
62. Lameira, A. R. *et al.* Proto-consonants were information-dense via identical bioacoustic tags to proto-vowels. *Nat. Hum. Behav.* **1**, 0044 (2017).
63. Lameira, A. R. *et al.* Orangutan information broadcast via consonant-like and vowel-like calls breaches mathematical models of linguistic evolution. *Biol. Lett.* **17**, 20210302 (2021).
64. Waser, P. M. & Brown, C. H. Habitat acoustics and primate communication. *Am. J. Primatol.* **10**, 135–154 (1986).
65. Waser, P. M. & Brown, C. H. Is there a 'sound window' for primate communication?. *Behav. Ecol. Sociobiol.* **15**, 73–76 (1984).
66. Brown, C. H., Gomez, R. & Waser, P. M. Old world monkey vocalizations: Adaptation to the local habitat?. *Anim. Behav.* **50**, 945–961 (1995).
67. Berglund, B., Hassmén, P. & Job, R. F. S. Sources and effects of low-frequency noise. *J. Acoust. Soc. Am.* **99**, 2985–3002 (1996).
68. Servedio, M. R. *et al.* Not just a theory—the utility of mathematical models in evolutionary biology. *PLOS Biol.* **12**, e1002017 (2014).
69. Wich, S. A. & Nunn, C. L. Do male 'long-distance calls' function in mate defense? A comparative study of long-distance calls in primates. *Behav. Ecol. Sociobiol.* **52**, 474–484 (2002).
70. Mehler, J., Peña, M., Nespor, M. & Bonatti, L. The, "Soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex* **42**, 846–854 (2006).
71. Vanormelingen, L., Faes, J. & Gillis, S. Language development in children from different SES backgrounds: Babbling onset and consonant characteristics. *Dutch J. Appl. Linguist.* **9**, 132–161 (2020).
72. Nazzi, T., Poltrock, S. & Von Holzen, K. The developmental origins of the consonant bias in lexical processing. *Curr. Dir. Psychol. Sci.* **25**, 291–296 (2016).
73. Toro, J. M., Nespor, M., Mehler, J. & Bonatti, L. L. Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychol. Sci.* **19**, 137–144 (2008).
74. Mourik, A. A., van Langevelde, F., van Tellinggen, E., Heitkönig, I. M. A. & Gaigher, I. Stability of wooded patches in a South African nutrient-poor grassland: Do nutrients, fire or herbivores limit their expansion?. *J. Trop. Ecol.* **23**, 529–537 (2007).
75. Ayers, A. M. *et al.* Illuminating movement? Nocturnal activity patterns in chacma baboons. *J. Zool.* **310**, 287–297 (2020).

## Acknowledgements

We thank Philip Faure for their assistance with fieldwork. We thank our two anonymous reviewers for constructive comments on previous versions of the manuscript. A.R.L. was supported by the UK Research and Innovation's Future Leaders Fellowship grant agreement number MR/T04229X/1. We thank two anonymous reviewers and editor Dr Metin Eren for their insightful comments.

## Author contributions

A.R.L.—conceptualisation; R.A.H. and A.R.L.—methodology; C.G.—formal analysis; C.G., R.A.H. and A.R.L.—investigation; A.R.L. and R.A.H.—resources; C.G.—data curation; C.G.—writing (original draft); C.G., R.A.H. and A.R.L.—writing (review and editing); A.R.L.—supervision; R.A.H.—project administration; A.R.L.—funding acquisition.



### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-48165-7>.

**Correspondence** and requests for materials should be addressed to C.G. or A.R.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023