Function and evolution of vibrato-like frequency modulation in mammals

Article (Accepted Version)

Function and evolution of vibrato-like frequency modulation in mammals

Benjamin D. Charlton[1], Anna M. Taylor[2] & David Reby.[3]

1: School of Biology and Environmental Science, Science Centre West, University College Dublin (UCD), Belfield, Dublin 4, Ireland

2: Bader International Study Centre, (Queen's University, Canada), Herstmonceux Castle, East Sussex, BN27 1RN, U.K.

3: Mammal Vocal Communication and Cognition Research Group, School of Psychology, University of Sussex, East Sussex, U.K.

Corresponding Author and Lead Contact: Benjamin D. Charlton (benjamin.charlton@ucd.ie)

SUMMARY

Why do distantly related mammals like sheep, giant pandas and fur seals produce bleats that are characterised by vibrato-like fundamental frequency (F0) modulation? To answer this question, we used psychoacoustic tests and comparative analyses to investigate whether this distinctive vocal feature has evolved to improve the perception of formants, key acoustic components of animal calls that encode important information about the caller's size and identity [1]**. Psychoacoustic tests on humans confirmed that vibrato-like F0 modulation improves the ability of listeners to detect differences in the formant patterns of synthetic bleat-like stimuli. Subsequent phylogenetically controlled comparative analyses revealed that vibrato-like F0 modulation has evolved independently in six mammalian orders in vocal signals with relatively high F0 and, therefore, low spectral density (i.e. less harmonic overtones). We also found that mammals modulate the vibrato in these calls over greater frequency extents when the number of harmonic overtones per formant is low, suggesting that this is a mechanism to improve formant perception in calls with low spectral density. Our findings constitute the first evidence that formant perception in non-speech sounds is improved by fundamental frequency modulation, and provide a mechanism for the convergent evolution of bleat-like calls in mammals. They also indicate that selection pressures for animals to transmit important information encoded by formant frequencies (on size and identity, for example) are likely to have been a key driver in the evolution of mammal vocal diversity.**

RESULTS AND DISCUSSION

According to the source-filter theory of voice production, the key components of mammal vocal signals are produced in two stages [2]. A source signal is produced in the larynx and characterised by its fundamental frequency (F0), which corresponds to the rate of vocal fold vibration in the larynx and determines the perceived pitch of the signal [3]. This source signal is then filtered in the supra-laryngeal vocal tract, whose resonance properties determine the formants that appear as broadband frequency maxima in the sound spectrum and determine the timbre of the signal [3]. One of the key assumptions of the source-filter theory is that F0 and formants can vary

independently, and both have been shown to provide receivers with important biosocial information on the caller's phenotype or motivational state [1]. However, the value of F0 can also affect the resolution of the formants and hence the availability of any information encoded by them. For example, when F0 is high, and the distance between the harmonic overtones (multiple integers of F0) is large, formant peaks are poorly resolved because the density of harmonics sampling the spectral envelope is relatively low [4] (Figure 1). Consistent with this, studies involving human listeners have confirmed that both the discrimination of vowels (which relies on formant perception) and the discrimination of formant patterns in synthetic voice-like signals are poorer when F0 is raised [5-7].

High spectral density can be achieved by a low F0 (Figure 1B) or broadband frequency noise/deterministic chaos [8]. However, vocal signals with low F0 or deterministic chaos may be selected against in affiliative contexts because they are typically associated with aggressive intent [9]. An alternative mechanism for increasing harmonic density in signals with relatively high F0s is to modulate F0 so that the harmonics scan a wider frequency bandwidth (Figure 1C). If F0 is repeatedly modulated then the likelihood and number of times that harmonics cross vocal tract resonances increase (Figure 1D), which should improve the resolution and perceptual salience of formants. However, work testing this hypothesis on humans has yielded conflicting results, with some studies indicating that vibrato-like F0 modulation may slightly improve formant perception [10-12], while others suggest that rapid F0 modulation hinders formant perception [13, 14]. Whether F0 modulation improves the salience of formants therefore remains an open question.

In this study we hypothesized that nonhuman animal vocalizations with sinusoidal F0 modulation (defined as an unbroken F0 contour that is modulated in a periodic, sinusoidal fashion), such as sheep bleats, horse whinnies and marmoset trills, may have evolved to highlight the formant pattern in high F0 calls with concomitant low harmonic density. To investigate this hypothesis, we combined human psychoacoustic tests and phylogenetically controlled comparative analyses. Psychoacoustic tests on human listeners sought to determine whether sinusoidal F0 modulation (hereafter SFM) improves the perceptual salience of formants in non-speech, bleat-like stimuli based on a source-filter production mechanism (for stimulus preparation see Supplementary Experimental Procedures). Participants were presented with synthetic vocal stimuli with or without SFM, and across a range of different mean

F0s. The stimuli were presented in pairs that were either identical, or differed only in the value of the fourth formant (F4). F4 was chosen because it is more likely to vary with vocal tract length and voice individuality than the lower three formants (which are used to produce vowel sounds), yet it remains well within the range of peak human auditory sensitivity [15]. To test whether humans can perceive subtle differences in formant patterns F4 was shifted up 3%, or down 3%, from its original value of 3850 Hz (Figure 2 and Figure S1). Previous work indicates that formant shifts of 3% and above are perceptible to humans whereas shifts below this magnitude a significantly harder to detect [7].

On the premise that F0 modulation increases the likelihood that harmonics cross formants (Figure 1D), we predicted that human participants would be best at detecting small shifts in the value of the fourth formant (F4) in stimuli with SFM (Figure 2A, and Figure S1). The best supported Generalised Estimating Equations (GEE) model, with the lowest Quasi Likelihood under Independence Model Criterion (QIC) value, included the mean F0 of the stimulus pairs, the subject's gender, and the interaction term 'F0 modulation condition x gender'. This binary logistic GEE model revealed a significant main effect of F0 modulation condition (SFM versus no F0 modulation) on classification performance (correct or incorrect) (GEE: N = 34, Wald $\chi^2_{1,\ 1223} = 28.32$, $p < 0.001$): subjects were better at classifying the stimuli with SFM as sounding the same or different than they were for stimuli with no F0 modulation (Figure 2C). Men were overall better than women at classifying stimuli (Wald $\chi^2_{1,\ 1223} = 6.66$, $p = 0.010$) and a significant interaction between F0 modulation condition and gender (Wald $\chi^2_{1,\ 1223} = 5.85$, $p = 0.016$) revealed that the positive effect of SFM on formant perception was strongest in male participants. Classification rates were also higher for stimuli with lower mean F0 (Wald $\chi^2_{3,\ 1221} = 45.29$, $p < 0.001$).

To further explore how SFM improves the perception of F4 shifts we calculated the number of times a harmonic entered the F4 bandwidth for the different classes of SFM stimuli (STAR Methods) and performed a linear regression between these values and the mean proportion of correct classifications. The number of times that a harmonic crossed into the F4 bandwidth in the SFM stimuli (F4 strikes) was positively correlated with the mean proportion of correct classifications ($F_{1,\ 9} = 5.58$, $p = 0.042$) (Figure 2D). Taken together, the results of the psychoacoustic tests confirm that SFM improves the ability of human listeners to perceive small shifts in the formant pattern of bleat-like

stimuli, and indicate that improved perception of formant variation in the SFM stimuli is driven by an increase in the number of times that harmonics strike formants.

Having established that SFM improves formant perception in human listeners, and making the assumption that these observations could generalise to other mammalian species, we then used phylogenetically controlled comparative analyses [16-18] to investigate the possible factors behind the evolution of SFM in nonhuman mammals. Data on mean F0 for 92 mammal species (Table S1) revealed that SFM has independently evolved in six mammalian orders: Carnivora, Primate, Artiodactyla, Perissondactyla, Rodentia, and Chiroptera (Figure 3). We predicted that SFM should evolve in nonhuman mammal calls with fewer harmonics per formant when compared to other call-types. A Brownian Motion + Pagel's lambda (λ) Phylogenetic Generalized Least Squares (PGLS) regression model with $\log_{10}$ body mass as a covariate (Table S2) was used to examine the relationship between harmonics to formant ratio (see STAR Methods for estimation of this parameter) and call-type (SFM calls versus all other call-types). SFM calls (N = 21 species) had a significantly lower $\log_{10}$ harmonics to formant ratio than other call-types (N = 71 species) (PGLS: estimate ± s.e. = -0.240 ± 0.094, $\lambda = 0.71$, $t_{3, 89} = -2.419$, $P = 0.013$) (Figure 4A).

Finally, we examined the relationship between the harmonics to formant ratio and percent modulation of F0 in SFM calls from 21 mammal species (Table S3). We predicted that F0 would be modulated over a greater frequency range in species that produce SFM calls with relatively lower harmonics to formant ratios, and in which the harmonics would be required to scan a larger frequency range in order to strike formants. A Brownian motion evolutionary model with habitat, social structure, and $\log_{10}$ body mass as covariates (Table S4) revealed that % F0 modulation was negatively correlated with $\log_{10}$ harmonics to formant ratio (estimate ± s.e. = -13.249 ± 5.185, $t_{5, 16} = -2.555$, $P = 0.021$) (Figure 4B). This result indicates that F0 is modulated over a relatively wider frequency range in SFM calls with fewer harmonics per formant, supporting the contention that SFM functions to scan the spectrum so that harmonics are more likely to excite formants.

The results of these psychoacoustic and comparative investigations provide insights into the convergent evolution of a distinctive form of mammal vocalization that is often referred to as a bleat or trill, and is characterized by periodic, sinusoidal F0 modulation (SFM). The psychoacoustic tests on human subjects confirmed that participants were significantly better at detecting small differences in the formant

pattern of synthetic vocal stimuli when those stimuli were characterized by SFM. We also provide clear evidence that formant perception in the SFM stimuli is improved as modulating harmonics cross formants to provide them with excitation energy and, consistent with previous findings [7], we found that men performed better than women in the discrimination task. Taken together, the results of the psychoacoustic tests support the general hypothesis that SFM enhances formant perception, and could therefore have evolved in nonhuman mammals for this purpose. Indeed, these observations are likely to generalize to receivers of nonhuman mammals, as formants have been shown to be both perceptually discriminable and important in size/identity communication in a wide range of mammalian species [1]

The subsequent comparative analysis revealed that SFM occurs in mammal calls characterized by a relatively higher F0 (and therefore a lower expected harmonic density in the absence of modulation) and SFM calls with fewer harmonics per formant tend to be modulated over greater frequency ranges. These results provide strong support for the hypothesis that SFM functions to highlight formants in calls with low harmonic densities, in which an unmodulated F0 would not produce sufficient spectral density to resolve the formants. Interestingly, a spectral density of less than 4.4 harmonics per formant is known to significantly impair human formant perception for vowel discrimination [4], and all the species producing calls with SFM had a ratio of harmonics to predicted formant spacing of 4.4 or less (Table S3). Accordingly, we suggest that once the harmonic density drops to around 4 harmonics per formant, SFM becomes an effective mechanism for highlighting functional information encoded by formants.

Mammal calls often have distinctive formant patterns, attributed to individual differences in vocal tract morphology, and several studies on humans and other nonhuman mammals have confirmed the importance of formants as cues to individual identity [1, 19-21]. We therefore suggest that SFM is important for highlighting individually distinctive formant patterns in a range of nonhuman mammal calls. Another mechanism for highlighting formants is to produce calls with a low F0 or broadband frequency noise [8]; however, because low frequency or noisy calls are typically associated with aggression [9], SFM is likely to be favoured for highlighting formant-related information in affiliative contexts. While the exact social context of SFM call production is not documented for all of the species in the analysis, only one of the 21 SFM call-types in the dataset is produced in an agonistic context (the giant

otter scream) (Table S3). All the other examples are produced in nonaggressive contexts when animals are thought to be promoting contact with other conspecifics and in which identity cueing is likely to be important, such as mother-offspring contact, promoting contact with mating partners, and reuniting with social group members. It is also possible that other characteristics of SFM calls are used for identity cueing in these contexts. For example, the rate and extent of F0 modulation contribute to individual vocal distinctiveness in Australian sea lions [22] and goats [23], respectively, and may therefore be individually distinctive components of SFM calls in other species.

In conclusion, these investigations have provided a highly plausible scenario for the convergent evolution of bleat-like calls in terrestrial mammal vocal signals, and highlight the importance of an interdisciplinary approach to tacking questions about the evolution of mammal vocal diversity. Future work should probe the ability of nonhuman mammals to discriminate between different callers using re-synthesised SFM calls with varying levels of F0 modulation. Investigations could also be extended to other vertebrates and to vocal signals that do not have SFM, but nevertheless contain strong F0 modulation that may function to highlight important formant-related information. By adopting a phylogenetically controlled comparative approach, these studies may reveal other examples of convergent evolution in vocal signal structure, and allow researchers to gain a better understanding of how and why certain features of animal vocalisations evolve independently in distantly related species.

Figure legends

Figure 1: The effect of different F0 characteristics on formant resolution. Panel A shows a power spectrum of a 500 Hz tone without F0 modulation and three formants at 550 Hz, 1650 Hz and 2750 Hz, labelled F1, F2 and F3, respectively. The power spectrum in panel B has a 250 Hz unmodulated F0 with the same formant pattern. The lower F0 and closer harmonic spacing provides greater formant resolution. In panel C a 500 Hz F0 is modulated in a sinusoidal fashion at 50% of its mean value. The SFM highlights the same formant pattern shown in panels A and B in more detail. Panel D shows a section of a giant panda bleat spectrogram in which the third and fourth harmonics (labelled upper and lower) of F0 strike the third formant. As a harmonic coincides with the centre frequency of the formant the darkening on the spectrogram

indicates an increase in frequency amplitude, which should increase the perceptual salience of the formant.

Figure 2: Psychoacoustic test stimuli and results. The upper panel shows spectrograms of vocal stimuli with (A) and without (B) sinusoidal F0 modulation (SFM). F0 = 250 Hz in both spectrograms and H1 and H2 refer to the first and second harmonic, respectively. The fourth formant (F4) is shifted up by 3% in the second presentation of each stimulus pair. Spectrogram settings: window length = 0.05 s, frequency step = 20 Hz, dynamic range = 40 dB, Gaussian window shape. Panel C shows the estimates ± SEM of the proportion of correct classifications made by participants for the two F0 modulation conditions. The relationship between the number of times harmonics enter the F4 bandwidth (F4 strikes) and the percentage of correct classifications in the SFM stimuli is shown in panel D. Each of the data points represent the mean correct classification rates across all 34 human subjects for each $\log_{10}$ transformed F4 strike value. See also Figure S1 for spectrograms of all the vocal stimuli.

Figure 3: Examples of sinusoidal F0 modulation (SFM) in mammals. Phylogeny used to control for shared ancestry between different mammal species (A) and examples of SFM calls from each of the 21 species that produce these vocalisations, often called "bleats" or "trills" (B). Red stars denote species that produce SFM calls. Spectrograms were generated using the following settings: window length = 0.01-0.05 s, frequency step = 20 Hz, dynamic range = 40 dB, Gaussian window shape. For the Baird's pocket gopher, pygmy marmoset, and grey-headed flying fox the spectrograms were taken from the literature [24-26] because we could not obtain recordings for these species. Panel C shows how the F0 characteristics of SFM calls were measured. Complete cycles of SFM are labeled FM1, FM2 etc. See also Table S1 and Table S3 for acoustic data on SFM calls and the sample composition.

Figure 4: Support for the hypothesis that sinusoidal F0 modulation (SFM) functions to highlight formants in calls with low spectral density. Panel A presents the estimates ± SEM of $\log_{10}$ harmonics to formants ratio for SFM (N = 21) and other call types (N = 71) derived from a Brownian Motion + Pagel's lambda (λ) model with $\log_{10}$ body mass as a covariate. SFM calls have significantly lower harmonics to

formants ratio than other call-types. Panel B shows the relationship between $\log_{10}$ harmonics to formants ratio and % F0 modulation in the SFM calls of 21 mammal species. The dotted line represents the slope and intercept obtained from the PGLS regression. % F0 modulation is significantly higher in SFM calls with lower harmonics to formant ratios. See also Table S1 and Table S3.

AUTHOR CONTRIBUTIONS

B.D.C. and D.R. conceived and designed the study; B.D.C. and D.R. wrote the manuscript; B.D.C. created the synthetic vocal stimuli, developed the software for capturing human responses to the stimuli, collated the data for the comparative analyses, and conducted the statistical analyses. A. T. performed the psychoacoustic experiments.

ACKNOWLEDGEMENTS

We would like to thank the students at the Bader International Study Centre for agreeing to participate in our study and Isabelle Charrier for providing recordings of Australian sea lions. We are also grateful to Kasia Pisanski for her very helpful comments on the manuscript.

REFERENCES

1. Taylor, A., Charlton, B.D., and Reby, D. (2016). Vocal production by terrestrial mammals: source, filter and function. In Vertebrate sound production and acoustic communication, R.A. Suthers, W.T. Fitch, R.R. Fay and A. Popper, eds. (Berlin: Springer International Publishing).
2. Fant, G. (1960). Acoustic Theory of Speech Production, (The Hague: Mouton).
3. Titze, I.R. (1994). Principles of Voice Production, (Englewood Cliffs, New Jersey: Prentice Hall).
4. Ryalls, J.H., and Lieberman, P. (1982). Fundamental-frequency and vowel perception. J. Acoust. Soc. Am. *72*, 1631-1634.
5. Assmann, P., and Nearey, T. (2008). Identification of frequency-shifted vowels. The Journal of the Acoustical Society of America *124*, 3203-3212.

6. Kewley-Port, D., Li, X., Zheng, Y., and Neel, A. (1996). Fundamental frequency effects on thresholds for vowel formant discrimination. The Journal of the Acoustical Society of America *100*, 2462-2470.
7. Charlton, B., Taylor, A., and Reby, D. (2013). Are men better than women at acoustic size judgements? Biology Letters *9*.
8. Fitch, W.T., and Hauser, M.D. (1995). Vocal production in nonhuman-primates - acoustics, physiology, and functional constraints on honest advertisement. Am. J. Primatol. *37*, 191-219.
9. Morton, E.S. (1977). On the occurrence and significance of motivation-structural rules in some birds and mammal sounds. Am. Nat. *111*, 855-869.
10. McAdams, S., and Rodet, X. (1988). The role of FM-induced AM in dynamic spectral profile analysis. In Basic Issues in Hearing, H. Duifhuis, Jorst, J. W., Witt, H. P., ed. (London: Academic Press), pp. 359-369.
11. Demany, L., and Semal, C. (1990). The effect of vibrato on the recognition of masked vowels. Percept. Psychophys. *48*, 436-444.
12. Erickson, M.L., and Gaskill, C.S. (2012). Can listeners hear how many singers are singing? The effect of listener's experience, vibrato, onset, and formant frequency on the perception of number of simultaneous singers. Journal of Voice *26*, 817.e811-817.e813.
13. Carlson, R., Fant, G., and Tatham, M.A.A. (1975). Two-formant models, pitch and vowel perception. In Auditory Analysis and Perception of Speech, G. Fant, Tatham, M. A. M., ed. (London: Academic Press), pp. 55-82.
14. Sundberg, J. (1977). Vibrato and vowel identification. Quarterly Progress and Status Report, 1-18.
15. Gelfand, S.A. (2011). Essentials of Audiology, (New York: Theime Medical Publishers).
16. Charlton, B.D., and Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. Nature Communications *7*, 12739.
17. García Navas, V., and Blumstein, D.T. (2016). The effect of body size and habitat on the evolution of alarm vocalizations in rodents. Biol. J. Linnean Soc. *118*, 745–751.
18. Bowling, D.L., Garcia, M., Dunn, J.C., Ruprecht, R., Stewart, A., Frommolt, K.H., and Fitch, W.T. (2017). Body size and vocalization in primates and carnivores. Sci Rep *7*, 41070.

19. Ghazanfar, A., Turesson, H., Maier, J., Vandinther, R., Patterson, R.D., and Logothetis, N. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. Current Biology *17*, 425-430.
20. Fitch, W.T., and Fritz, J.B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalisations. J. Acoust. Soc. Am. *120*, 2132-2141.
21. Rendall, D., Owren, M.J., and Rodman, P.S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. J. Acoust. Soc. Am. *103*, 602-614.
22. Charrier, I., and Harcourt, R. (2006). Individual vocal identity in mother and pup Australian sea lions (*Neophoca cinerea*). J. Mammal. *87*, 929-938.
23. Briefer, E., and McElligott, A.G. (2011). Mutual mother offspring vocal recognition in an ungulate hider species (*Capra hircus*). Animal Cognition *14*, 585-598.
24. De La Torre, S., and Snowdon, C.T. (2009). Dialects in pygmy marmosets? Population variation in call structure. Am. J. Primatol. *71*, 333-342.
25. Devries, M.S., and Sikes, R.S. (2008). Vocalisations of a North American subterranean rodent *Geomys breviceps*. Bioacoustics *18*, 1-15.
26. Nelson, J.E. (1964). Vocal Communication in Australian Flying Foxes (*Pteropodidae*; *Megachiroptera*). Ethology *21*, 857-870.
27. Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W., Boakes, E.H., Carbone, C., et al. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. In Ecology, Volume 90. (Ecological Society of America), pp. 2648-2648.
28. Charlton, B. (2014). Vocal distinctiveness in the harsh coughs of southern hairy-nosed wombats (*Lasiorhinus latifrons*). Acta Acustica United With Acustica *100*, 719-723.
29. Hauser, M.D. (1993). The evolution of nonhuman primate vocalisations: effects of phylogeny, body weight and social context. Am. Nat. *142*, 528-542.
30. Mitani, J.C., and Stuht, J. (1998). The evolution of nonhuman primate loud calls: Acoustic adaptation for long-distance transmission. Primates *39*, 171-182.
31. Reby, D., and McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. Anim. Behav. *65*, 519-530.

32. Charlton, B., Zhihe, Z., and Snyder, R. (2009). The information content of giant panda, Ailuropoda melanoleuca, bleats: acoustic cues to sex, age and size. Anim. Behav. *78*, 893-898.
33. Vannoni, E., and McElligott, A.G. (2007). Individual acoustic variation in fallow deer (*Dama dama*) common and harsh groans: a source-filter theory perspective. Ethology *113*, 223-234.
34. Charlton, B., Ellis, W., McKinnon, A., Cowin, G., Brumm, J., Nilsson, K., and Fitch, W. (2011). Cues to body size in the formant spacing of male koala (Phascolarctos cinereus) bellows: honesty in an exaggerated trait. J. Exp. Biol. *214*, 3414-3422.
35. Cao, D., Zhou, H., Wei, W., and Lei, M. (2016). Vocal repertoire of adult captive red pandas (*Ailurus fulgens*). Anim. Biol. *66*, 145-155.
36. Charlton, B., Reby, D., Ellis, W., Brumm, J., and Fitch, W. (2012). Estimating the Active Space of Male Koala Bellows: Propagation of Cues to Size and Identity in a Eucalyptus Forest. Plos One *7*.
37. Wiley, R.H., and Richards, D.G. (1978). Physical constraints on acoustic communication in atmosphere - implications for evolution of animal vocalizations. Behav. Ecol. Sociobiol. *3*, 69-94.
38. Boersma, P., and Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. . 6.0.29 Edition.
39. (R Core Team 2015). A language and environment for statistical computing (R Foundation for Statistical Computing).
40. Højsgaard, S., Halekoh, U., and Yan, J. (2016). geepack: Generalized Estimating Equation Package. R package version 1.2-1.
41. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., and Van Willigen, B. (2017). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131.
42. Morrow, E.H., and Fricke, C. (2004). Sexual selection and the risk of extinction in mammals. Proc Biol Sci *271*, 2395-2401.
43. Harvey, P.H., and Pagel, M.D. (1991). The comparative method in evolutionary biology, (Oxford University Press).
44. Pan, W. (2001). Akaike's information criterion in generalized estimating equations. Biometrics *57*, 120-125.

45. Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., Macphee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., and Purvis, A. (2007). The delayed rise of present-day mammals. Nature *446*, 507-512.

46. Bartoń, K. (2016). MuMIn: Multi-Model Inference. R package version 1.15.6.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for protocols and datasets should be directed to and will be fulfilled by the Lead Contact, Benjamin D. Charlton (benjamin.charlton@ucd.ie).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A total of 34 college undergraduates (15 males and 19 females) aged between 18-22 years (mean age = 18.6 years) completed the psychoacoustic experiment at the Bader International Study Centre. All the participants gave informed consent. The University College Dublin Office of Research Ethics approved the human psychoacoustic tests (LS-15-39-Charlton).

METHOD DETAILS

Creation of vocal stimuli for the psychoacoustic tests

A series of vocal stimuli with and without F0 modulation were synthesized using Praat (www.praat.org) and following the principles of the source-filter theory of voice production [2, 3] (Figure S1). The stimuli consisted of a two second long harmonic tone, the 'source', combined with a formantGrid pattern, the 'filter'. The harmonic tones were created from a sine wave (using "Create Sound from formula... " in Praat) and the following settings: Channels: Mono, Start time (s) = 0, End time (s) = 2, Sampling frequency (Hz) = 44100, Formula: sin(2*pi*'f0'*x + ('f0modupdown'/10)*(sin(2*pi*('f0modrate'*x)))). The parameters 'f0' and 'f0modupdown' in the formula were adjusted to create sine waves with 4 different mean F0s (250 Hz, 500 Hz, 750 Hz, and 1000 Hz) and 3 levels of F0 modulation: no modulation, SFM at 5 cycles per second, and SFM at 10 cycles per second. The variation in mean F0 and the rate of F0 modulation in the SFM generated variability in the amount of times that harmonics entered the F4 bandwidth, which in turn allowed us

to examine how this factor affected classification performance. The starting phase for all the sinusoidal tones was set at zero degrees and the extent of sinusoidal F0 modulation ('f0modupdown') was fixed at 100Hz. The modulated sine waves were then converted into harmonic sounds using the 'To Sound (phonation)…' command in Praat and the standard settings (Sampling frequency (Hz) = 44100, Adaptation factor = 1.0, Maximum period (s) = 0.05, Open phase = 0.7, Collision phase = 0.03, Power 1 = 3.0, Power 2 = 4.0, no 'Hum').

The formant pattern was created using a formantGrid in Praat with equally spaced formants ("schwa") to approximate an idealized uniform straight tube, or an unperturbed vocal tract. The formantGrid consisted of 10 formants at: F1 = 550, F2 =1650, F3 = 2750, F4 = 3850, F5 = 4950, F6 = 6050, F7 = 7150, F8 = 8250, F9 = 9350 and F10 = 10450 Hz. The default formant bandwidth was used so that F1 had a bandwidth of 60Hz, and subsequent formants in adjacent tiers had the bandwidth increased by 50 Hz throughout the time domain; so that F2 bandwidth = 110 Hz, F3 bandwidth = 160 Hz, F4 bandwidth = 210 Hz, and so on. The overall ΔF of 1100 Hz corresponds to a vocal tract length of 15.9 cm, which falls within the typical human range [3]. To create our different formant patterns the baseline formantGrid then had the frequency value of the 4th formant (F4) shifted up 3%, or down 3%, from its original value of 3850 Hz, using the 'Modify Formula (frequencies)…' command, to create a total of three different formantGrid patterns: F4 shifted up 3%, down 3%, and no shift. Previous work indicates formant shifts less than 3% are significantly harder for humans to perceive than shifts of this magnitude or above [7]. The fourth formant was chosen because it sits in the range of peak human auditory sensitivity [15]. The harmonic tones and formantGrid patterns were combined to create the synthetic vocal stimuli using the 'Filter' option in Praat, the intensity of all the stimuli was standardized to 65 dB, and a 0.1 s fade out was applied. For each of the four mean F0s and three F0 modulation conditions (no modulation, SFM at 5 cycles per second, and SFM at 10 cycles per second) the stimuli were arranged in matched pairs so that stimuli with the original formant pattern (baseline condition) were followed 0.5 seconds either by exactly the same stimuli, or stimuli with a formant pattern that had F4 shifted up or down by 3% (Figure 2A and B). In this way a set of 36 different stimuli were created (Figure S1): 3 F0 modulation conditions x 4 mean F0 conditions x 3 shift directions. The sequences were saved as WAV files at 44.1 kHz and 16 bits amplitude resolution.

Calculation of F4 strikes in the vocal stimuli

The F4 bandwidths for the three different formant patterns were 3630-3840 Hz, 3745-3955 Hz and 3861-4071 Hz (calculated as a 210 Hz bandwidth for F4 shifted down 3%, not shifted, and shifted up 3% from its original value of 3850). We then calculated the frequency range that each harmonic in the stimuli would scan over one complete cycle of SFM, and noted if harmonics crossed into the F4 bandwidths for each F4 shift condition over the four stimulus mean F0s. To calculate the number of F4 strikes per second we then multiplied the number of harmonics entering the F4 bandwidth by 5 for SFM stimuli modulated at 5 cycles per second and 10 for SFM stimuli modulated at 10 cycles per second.

Experimental procedure for the psychoacoustic tests

Participants were seated in a quiet room for the duration of the experiment and the stimuli were presented using a Dell Inspiron 1750 computer with the volume setting at a comfortable pre-set level. All participants wore Dynamode DH-660MV headphones for the stimulus presentation. The experimenter provided detailed instructions prior to the experiment and remained in the room, but out of sight of the participant throughout the experimental procedure. Participants were informed that they would hear pairs of computer-generated audio stimuli, that the two repetitions might be identical to one another in frequency or slightly different, and that their task was to rate whether each pair of audio stimuli sound the same or different by clicking on the appropriate button on the computer screen bearing that label ('Different' or 'Same'). Each participant received the 36 unique stimulus pairs representing three F0 modulation conditions at four levels of mean F0 with F4 shifted up or down 3%, or not shifted (3 F0 modulation conditions x 4 mean F0 conditions x 3 F4 shifts). Custom-written software in Python v2.8 was used to randomize stimulus presentation and collect responses.

Data sources for the comparative analyses

For the comparative analyses we collated acoustic and body weight data from published studies (Table S1). Whenever possible acoustic and body mass data were obtained from the same published source. If body weight data were not available from the acoustic studies we referred to the PANTHERIA v.1 database [27]. We collected acoustic data on mean F0 from 92 species and mean formant frequency spacing (ΔF) values from 23 species (Table S1). F0 and formant data were available for 22 species

(southern hairy-nosed wombat harsh coughs have published formant data but do not have an observable F0 [28]). Accordingly, we estimated ΔF for the remaining 70 species using a model (Figure S2). For four species mean F0 was taken as the average of the minimum and maximum reported values [sensu 29, 30]. For the Diana monkey, banded mongoose and northern fur seal mean F0 was measured from a published spectrogram using a clear ruler to extrapolate from the axes. To calculate ΔF we used the first 2 to 9 formant frequencies (mean = 5) and the regression method of Reby & McComb [31], in which the formant frequency values are plotted against those that would be expected if the vocal tract was a straight uniform tube closed at one end (the glottis) and open at the other (the mouth). This method is a very accurate way to estimate ΔF in species with unevenly spaced formants [as is typical in mammals 31, 32-34]. We excluded species with adaptations that allow them to produce exceptionally low or high frequencies (F0 or formants) for their size (such as descended larynges, hypertrophied larynges, nasal proboscis's, and additional resonators, for example), and restricted the dataset to adult mammals. To find examples of calls with sinusoidal F0 modulation (SFM) we examined spectrograms from published papers. SFM is defined as an unbroken F0 contour that is modulated in a periodic, sinusoidal fashion (see Figure 3 for examples). In total we were able to identify SFM vocalisations for 21 mammal species (Figure 3, Table S3). Acoustic data for SFM characteristics were only reported for eight taxa (grey-headed flying fox, giant panda, goat, pygmy marmoset, South American fur seal, Australian sea lion, grey mouse lemur, and Baird's pocket gopher) (Table S3). Thus, to determine whether F0 modulation extent increases as the ratio of harmonics to formants decreases, we obtained recordings of SFM vocalisations for 13 additional species for which we could not find the necessary acoustic data from the Animal Sound Archive at the Museum für Naturkunde Berlin (http://www.animalsoundarchive.org/) and the Macaulay Library (http://macaulaylibrary.org/) (see Table S3 for sample composition). Red panda bleats were obtained from Cao et al supplementary material [35] and a Hollywood Edge (Animal Trax) audio CD, and capybara whines were downloaded from the "capybaraworld" website (https://capybaraworld.wordpress.com/) (Table S3). Because the physical and social environment shapes the acoustic features of vocal signals [30, 36, 37], we also collected data on the typical habitat (terrestrial or arboreal) and social structure (social versus solitary) for each of the species in our comparative analyses

from the Encyclopaedia of Life website (http://eol.org/) to control for these factors in the analyses.

Calculation of harmonics to formant ratio

To obtain the harmonics to formants ratio for all the species in our phylogenetic analyses it was necessary to have data on ΔF. We were able to obtain data on F0 and formant frequencies for 22 taxa without adaptations to lower formants (such as descended larynges and additional resonators, for example) and used this to examine the relationship between $\log_{10}$ ΔF and $\log_{10}$ body mass. An Ordinary Least Squares (OLS) model without species' habitat (terrestrial or arboreal) and social structure (solitary or social) as covariates proved to be the best supported model, with the lowest AICc value (Table S2). This model revealed a very close negative relationship between $\log_{10}$ ΔF and $\log_{10}$ body mass ($R^2 = 0.88$, estimate ± s.e. = -0.295 ± 0.024, $t_{1,\ 21} = -12.49$, $P < 0.001$) (Figure S2), and subsequently allowed us to predict ΔF for the 70 species in our analysis for which we could not obtain published data on formant frequencies (Table S1). Log_{10} ΔF was calculated by substituting "x" for a given specie's $\log_{10}$ body mass in the OLS model regression equation: (-0.294725x + 4.292901). We were then able to calculate the inverse log (10^x) of the Y value obtained, and predict harmonics to formant ratio by dividing predΔF (the predicted average spacing between the formants) by mean F0 (the spacing between harmonics).

Acoustic analysis of SFM calls

To measure mean F0 and the extent of F0 modulation in the SFM vocalisations we used custom-built scripts in Praat 5.1.32 [38] (www.praat.org). A cross-correlation algorithm was used to produce time-varying numerical representations of the F0 contour for each call, and we also viewed spectrograms corresponding to the individual calls to check the acoustic values using on screen cursors. The time step in the analysis was 0.01, and a five-point average smoothing filter was used to remove any rapid variations caused by analysis imprecision. To limit the possibility of "octave jumps," the minimum and maximum values for F0 were set according to the F0 contour as observed on the spectrogram. F0 modulation extent for each call was calculated as the mean peak-to-peak variation of each cycle of SFM in hertz divided by the mean F0 [32] (Figure 3C). These values were then multiplied by 100 to give mean % F0 modulation

for each of the calls in the analysis, thereby standardising the extent of F0 modulation across species and calls.

QUANTIFICATION AND STATISTICAL ANALYSIS

All the statistics were conducted using R v3.2.3 [39], two-tailed probability values are quoted, and significance levels were set at $p = 0.05$. To examine variation in listeners' classification performance for the different classes of psychoacoustic stimuli we ran generalized estimating equations (GEE) using the geepack package in R [40]. For the comparative analyses we conducted phylogenetic generalised least squares (PGLS) regressions using the using the gls function (nlme package) in R [41]. The use of PGLS regressions allowed us to control for the confounding effects of shared phylogenetic ancestry [42, 43].

Psychoacoustic tests

A binomial GEE with a logit link function was used to examine variation in classification performance with the participants' responses entered as a binary logistic (correct or incorrect) dependent variable. F0 modulation condition (no F0 modulation versus SFM) and mean F0 condition (250, 500, 750 and 1000 Hz) were entered into the GEE as within-subjects factors, and gender as a between-subjects factor. The best supported model with the lowest quasi-likelihood under the independence model criterion (QIC) [44] included the interaction term F0 modulation condition x gender.

Comparative analyses

The PGLS regressions used untransformed branch lengths and splitting dates from a recent molecular phylogeny of mammals [45] and the maximum-likelihood method to estimate Pagel's lamda (λ). The first PGLS model allowed us to determine the relationship between $\log_{10}$ ΔF and $\log_{10}$ body mass (in grams) while controlling for habitat (arboreal versus terrestrial), social structure (solitary versus social) and phylogeny (Figure S2, Table S5). We were then able to predict ΔF for 70 species without published formant frequency data using the regression equation from this PGLS model. Two subsequent PGLS regressions were then used to test whether SFM calls have lower harmonics to formant ratios and whether % modulation of mean F0 in SFM calls increases as the harmonics to formant ratio decreases. For each hypothesis

we computed five PGLS regression models that were designed to test a different evolutionary scenario, and chose the most parsimonious model with the lowest Akaike Information Criterion statistic corrected for sample size (AICc). The different models were an Ornstein–Uhlenbeck (OU) model of evolution, a non-phylogenetic ordinary least-squares (OLS) model, a pure Brownian motion (BM) model, and two restricted maximum-likelihood (REML) Brownian motion models that allow parameters to vary with the strength of the phylogenetic signal, a Brownian motion + Pagel's lambda (BM + λ) and a Brownian motion + Grafen's rho (BM + ρ) model [for more details see 16]. The 'dredge' function in R (MuMIn' package [46]) was used to iterate through all variable combinations in a global model that included $\log_{10}$ body mass, habitat, and social structure, so we could select the best supported model with the lowest AICc value (see Table S2, Table S4, Table S5 for model outputs). Harmonic density was $\log_{10}$ transformed to achieve a normal distribution.

DATA AND SOFTWARE AVAILABILITY

The datasets necessary to run the analyses included in this paper are provided at: https://data.mendeley.com/datasets/6yrf6xjjp6/1