Recursive self-embedded vocal motifs in wild orangutans

Adriano R. Lameira[1*], Madeleine E. Hardus[^], Andrea Ravignani[2,3], Teresa Raimondi[4], Marco Gamba[4]

[1]Department of Psychology, University of Warwick, Coventry, UK
[2]Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[3]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Denmark
[4]Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy
[^]Independent researcher
*corresponding author: adriano.lameira@warwick.ac.uk

Abstract

Recursive procedures that allow placing a vocal signal inside another of similar kind provide a neuro-computational blueprint for syntax and phonology in spoken language and human song. There are, however, no known vocal patterns among nonhuman primates arranged in self-embedded combinations that evince vocal recursion or potential insipient forms and neuro-procedures thereof, suggesting a neuro-cognitive transformation exclusive to humans. Here, we uncover that wild flanged male orangutan long calls show two hierarchical strata, wherein rhythmically isochronous call sequences are nested within self-similar isochronous call sequences. Remarkably, three unrelated recursive motifs occurred simultaneously in long calls, refuting that motifs resulted from three parallel linear procedures or that motifs were simple anatomical artifacts of bodily constrains. Findings represent a case of recursive hominid vocal production in the absence syntax, semantics, phonology or music. Second-order combinatorics, 'sequences within sequences', involving hierarchically organized and cyclically structured vocal sounds in ancient hominids may have preluded the evolution of recursion in modern language-able humans.

Introduction

Among the many definitions of recursion (Martins, 2012), the view that it represents the capacity to iterate a signal within a self-similar signal has crossed centuries and disciplines, from von Humboldt (1836) and Hockett (1960), to Mandelbrot (1980) and Chomsky (2010); from fractals in mathematics (Mandelbrot, 1980) to generative grammars in linguistics (Chomsky, 2010). Across varying terminologies, the common denominator across fields is that to re-curse (that is, to 're-*invoke*') or re-iterate is an algorithmic "hack" to produce infinite signal states from a finite signal set. Although classically associated with syntax (Chomsky, 2010; Idsardi et al., 2018), recursive signal production and the resulting self-embedded structures have also been recognised in phonology (Bennett, 2018; Elfner, 2015; Kabak and Revithiadou, 2009; Nasukawa, 2015, 2020; Vogel, 2012) and (verbal and non-verbal) music (Jackendoff, 2009; Koelsch et al., 2013; Martins et al., 2017; Sharma and Chimalakonda, 2018). Given that language and music are uniquely human, this hints at a neuro-cognitive or neuro-computational transformation in the human brain that enabled the emergence of multiple open-ended communication systems in the human lineage, but seemingly, none other (Hauser et al., 2002). The absence of vocal structures in (nonhuman) primates that could help inform insipient or transitional recursive states, namely within the hominid family, has led some scholars to question altogether whether recursion was the result of natural selection (Berwick and Chomsky, 2019; Bolhuis and Wynne, 2009) and to rebut the role of shared ancestry and evolution as an incremental path-dependent process (cf. de Boer et al., 2020; Jackendoff and Pinker, 2005; Kershenbaum et al., 2014; Martins and Boeckx, 2019), favouring instead sudden "hopeful monster" mutant scenarios.

Decades-long debates on the evolution of recursion have ensued, carved around the successes and limitations of empirical comparative animal research. For example, perception and processing of syntax-like vocal combinatorics has been identified in some bird (Engesser et al., 2019, 2016; Gentner et al., 2006; Liao et al., 2022; Suzuki et al., 2016, 2017) and primate species (Jiang et al., 2018; Wang et al., 2015; Watson et al., 2020) but results' interpretation has received various criticisms (Johan J Bolhuis et al., 2018; Bowling and Fitch, 2015; Corballis and Corballis, 2014; Rawski et al., 2021). For example, these animal studies have, thus far, almost exclusively focused on *perception* (cf. Ferrigno et al., 2020; Liao et al., 2022): animals recognizing or responding to recursive signal structures; however, processing operations of recursive signals are not necessarily recursive themselves, for example (Corballis and Corballis, 2014; Miyagawa, 2021). Nonetheless, critically, most research has focused on laboratory animals in artificial test settings and/or who were presented with synthetic stimuli after dedicated human training

(Gentner et al., 2006; Jiang et al., 2018; Liao et al., 2022; Watson et al., 2020), including the few exceptions based on production instead of perception (Ferrigno et al., 2020; Liao et al., 2022). This opens comparative research to two vital drawbacks. First, results are mute about evolutionary precursors and processes. Settings and stimuli have not been those to which species adapted over evolutionary time, with most species tested thus far being too distantly related to humans to allow inferences about recursion emergence among hominids in the first place. That animals may be able to recognise recursive structures or engage in non-vocal recursive action after undergoing fixed training protocols with human-designed stimuli will never per se indicate whether recursion has been selected in the wild, including human ancestors.

Second, results have been slanted by the dominant classic theory that, counter-intuitively, disfavours a gradual evolutionary scenario for recursion and instead conceives it as a punctuated event (Berwick and Chomsky, 2019; Bolhuis et al., 2015; Bolhuis and Wynne, 2009). Namely, experimental stimuli have consisted of artificial recursive signal sequences organized along a single temporal scale (though not structurally linear), similarly with how *Merge* and syntax operate. However, recursive signal structures can also unfold in other manners, such as across nested temporal scales and in the absence of semantics (Fitch, 2017a), as in music. But these remain utterly unprobed and untested.

An approach to the production of recursive signals, complementary to perceptual studies, is thus, desirable. Data from *wild* primates in particular may help infer signal patterns that were recursive in some degree or kind in an extinct past and subsequently moulded into the recursive structures observed today in modern humans. By virtue of their own primitive nature, proto-recursive structures did not likely fall within modern-day classifications, and thus, will often fail to be predicted based on assumptions guided by full-fledged language (Kershenbaum et al., 2014; Miyagawa, 2021). To this end, a structural approach is particularly advantageous. First, no prior assumptions are required about species' cognitive capacities. These are directly inferred from how signal sequences are organised. For example, Chomsky's definition of recursion (Chomsky, 2010) can generate non-self-embedded signal structures, but these would be operationally undetectable amongst other signal combinations. Second, no prior assumptions are required about signal meaning. There are no certain parallels with semantic content and word meaning in animals, but analyses of signal patterning allows to identify similarities between non-semantic (nonhuman) and semantic (human) combinatoric systems (Lipkind et al., 2013; Sainburg et al., 2019). Therefore, the search for recursion can be made in the absence of meaning-base operations, such as *Merge*, and more generally, semantics and syntax. Third, no prior assumptions are required about signal function. Under punctuated or gradual evolutionary

hypotheses, ancestral signal function (whether cooperative, competitive or otherwise) is expected to have derived or been leveraged by its proto-recursive structure. Otherwise, once present and despite its origin process, recursion would not have been fixated among human ancestral populations. A structural approach opens, therefore, the field to untapped signal diversity in nature and yet unrecognised bona fide combinatoric states *within the human clade*.

Here, undertaking an explorative structural approach to recursion, we provide evidence for recursive self-embedded vocal patterning in a (nonhuman) great ape, namely, in the long calls of flanged orangutan males in the wild. We conducted precise rhythm analyses (De Gregorio et al., 2021; Roeske et al., 2020) of 66 long call audio recordings produced by 10 orangutans (*Pongo pygmaeus wurmbii*) across approximately 2510 observation hours at Tuanan, Central Kalimantan, Indonesian Borneo. We identified 5 different element types that comprise the structural building blocks of long calls in the wild (Hardus et al., 2009; Lameira and Wich, 2008), of which the primary type are full pulses (Fig. 1A). Full pulses do not, however, always exhibit uninterrupted vocal production throughout a long call [as during a long call's climax (Spillmann et al., 2010)] but can break-up into 4 different "sub-pulse" element types: (*i*) grumble sub-pulses [quick succession of staccato calls that typically constitute the first build-up pulses of long calls (Hardus et al., 2009)], (*ii*) sub-pulse transitory elements and (*iii*) pulse bodies (typically constituting pulses before and/or after climax pulses) and (*iv*) bubble sub-pulses (quick succession of staccato calls that typically constitute the last tail-off pulses of long calls) (Fig. 1A). We characterised long calls' full- and sub-pulses' rhythmicity to determine if orangutan long calls present a re-iterated structure across different hierarchical strata. We extracted inter-onset-intervals (IOIs; i.e. time difference between the start of a vocal element and the preceding one - t_k) from 8993 vocal long call elements (Fig. 1A): 1930 full pulses (1916 after filtering for $0.025<tk<5s$), 757 grumble sub-pulses (731), 1068 sub-pulse transitory elements (374), 816 pulse bodies (11) and 4422 bubble sub-pulses (4193). From the extracted IOIs, we calculated their rhythmic ratio by dividing each IOI by its duration plus the duration of the following interval. We then computed the distribution of these ratios to ascertain whether the rhythm of long call full and sub-pulses presented natural categories, following published protocols (De Gregorio et al., 2021; Roeske et al., 2020) (Fig. 1B, C, D).

Results

The density probability function of orangutan full pulses showed one peak (r_k=0.493) in close vicinity to a theoretically pure isochronic rhythm, that is, full pulses were regularly paced at 1:1 ratio, following a constant tempo along the long call (Fig. 1C). Our model (GLMM, *full* model vs

null model: Chisq=298.2876, df=7, p<0.001; see Supplementary Materials) showed that pulse type, range of the curve (on-off-isochrony), and their interaction, had a significant effect on the count of r_k values. In particular, full pulses' isochronous peak tested significant (t.ratio=-15.957, p<0.0001), that is, the number of r_k values falling inside on-isochrony range was significantly higher than the number of r_ks falling inside the off-isochrony range (Fig. 1C). Critically, three (of the four) orangutan sub-pulse element types – grumble sub-pulses, sub-pulse transitory elements and bubble sub-pulses – also showed significant peaks (grumble sub-pulses: t.ratio = -5.940, p<.0001; sub-pulse transitory elements: t.ratio=-4.048, p=0.0001; bubble sub-pulses: t.ratio= -10.640, p<.0001) around pure isochrony (peak r_k: grumble sub-pulses = 0.501; sub-pulse transitory elements=0.495; bubble sub-pulses=0.502; Fig. 1C). That is, sub-pulses were *regularly paced within regularly paced* full pulses, denoting isochrony within isochrony (Fig. 1C) at different average tempi (mean t_k (sd): full pulses=1.696 (0.508); grumble sub-pulses=0.118 (0.111); sub-pulse transitory elements=0.239 (0.468); bubble sub-pulses= 0.186 (0.292); Fig. 1B). Overall, sub-pulses' t_k was equivalent to 0.046 of their comprising full-pulses (Fig. 1D), which put sub-pulses at an approximate ratio of 1:20 relative to that of full-pulses, the smallest categorical temporal rhythmic interval registered thus far in a vertebrate (De Gregorio et al., 2021; Roeske et al., 2020). Permuted discriminant function analyses (Mundry and Sommer, 2007) (crossed, in order to control for individual variation) in R (Team, 2013) based on seven acoustic measures extracted from grumble, transitory and bubble sub-pulses confirmed that these represented indeed distinct sub-pulse categories, where the percentage of correctly classified selected cases (62.7%) was significantly higher (p=0.001) than expected (37%).

Discussion

Orangutan long call nested rhythmic patterns reveal self-similar embedded isochrony in the vocal production of a wild great ape, notably, with two discernible structural strata – the full- and sub-pulse level – and three non-exclusive rhythmic arrangements in the form of [isochrony^{A} [isochrony^{a,b,c}]].

Human and nonhuman great apes have similar auditory capacities (Quam et al., 2015). There are no identified skeletal differences in inner ear anatomy that could suggest significantly distinct sound sensitivity, resolution or activation thresholds in the time domain (Quam et al., 2015). Humans perceive an acoustic pulse as a continuous pitch, instead of a rhythm, at rates higher than 30 Hz (i.e., 30 beats per second). Long call sub-pulses exhibited average rhythms at ~9.263 (3.994) Hz [i.e., t_k=0.184 (0.303)s]. Therefore, hominid ear anatomy offers strong confidence that orangutans, like humans (and other great apes) perceive sub-pulse rhythmic

motifs as such, i.e., as a train of signals, instead of one uninterrupted signal. Assuming otherwise would imply that auditory time-resolution differ by more than one order of magnitude between humans and other great apes in the absence of obvious anatomical culprits.

The simultaneous occurrence of non-exclusive recursive patterns excludes the likelihood that orangutans concatenate long calls and their subunits in linear structure without any recursive processes. To generate the observed vocal motifs as linear structure, three independent neuro-computational procedures would need to run in parallel to generate distinct isochronic rhythms at the sub-pulse level, whilst being indistinguishable, transposable and/or interchangeable at the pulse level without interference, which would be unlikely, if theoretically possible at all. Non-exclusive simultaneous recursive patterns also help exclude the probability that recursion was the primary by-product of anatomic constrains, such as breath length, heart beat and other physiological rhythmic processes or movements (Pouw et al., 2020). Such processes could generate various simultaneous rhythmic patterns; however, these would be expected to be in the form of harmonics of the same base "carrier" rhythm. Yet, the three observed rhythmic arrangements at the sub-pulse level were not related to the pulse level by any small integer ratios. Together, this strongly suggests that the observed recursive self-embedded motifs are most likely generated by a recursive neurological procedure or recursion algorithm.

Recursive self-embedded vocal motifs in orangutans indicate that vocal recursion among hominids is not exclusive to human vocal and cognitive systems. This is not to suggest that they exhibit *all* properties that recursion exhibits in modern language-able humans. Such expectation would be unreasonable, as it would imply that no evolution occurred in >10 million years since the split between the orangutan and human phylogenetic lineages. Thus, any apparent differences with recursion in today's syntax, phonology, or music do not invalidate the probability that the reported vocal motifs represent an ancient, or potential ancestral, state for the ensuing evolution of recursion along the human clade.

Recursion and fractal phenomena are prevalent across the universe. From celestial and planetary movement to the splitting of tree branches and river deltas, and the morphology of bacteria colonies; Patterns within self-similar patterns are the norm, not the exception. This makes the seeming singularity of human recursion amongst vertebrate signals only the more enigmatic. Our findings indicate that ancient vocal patterns organized across nested structural strata were likely present in ancestral hominids. Recursive vocal production likely predated the evolutionary emergence of (spoken) language within the Hominid family and along the human lineage. This data-driven possibility poses new evolutionary trajectories and timelines distinct from the classic theoretical notions that the advent of recursion was a saltational all-or-nothing

event that took place only recently in modern humans (Berwick and Chomsky, 2019). Gradual cumulative evolutionary scenarios for the emergence and evolution of recursion (de Boer et al., 2020; Martins and Boeckx, 2019) can and should be considered and investigated, with the advantage that living primate models in general, and natural hominid behaviour in particular, offer superior empirical and comparative validity beyond purely theoretical considerations (Lameira et al., 2021). A combination of both approaches will likely pay the highest heuristic dividends. For example, future research may implement playback experiments designed to clarify whether or how great apes cognitively represent natural recursive self-embedded motifs as informed by previous designs and theoretical considerations (Corballis and Corballis, 2014; Engesser et al., 2016; Watson et al., 2020).

Implications for the evolution of recursion and cognition

The presence of recursive self-embedded vocal motifs in orangutans carries four major implications for the evolution of recursion and cognition. First, much ink has been laid on the topic, yet, the possibility of self-embedded isochrony, or alternatively, non-exclusive self-embedded patterns occurring within the same signal sequence, has on no account been formulated or conjectured as a possible state of recursive signalling, be it in vertebrates, mammals, primates or otherwise, extant or extinct. This suggests that controversy may have thus far been underscored by data-poor circumstances, namely, lack of rigorous and detailed knowledge about call combinations in wild primates in general, and great apes in particular. The presence of recursive vocal patterns in a wild great ape in the absence of syntax, phonology or music helps prevent prior assumptions that recursion precursors ought, or should be expected to, operate as modern-day recursion. Orangutan recursive vocal patterns provide the first data point and open a new charter for possible insipient or transitional states within the hominid family that, whilst inherently distinct from modern-day recursion, were nonetheless homologous to it and fully functional in their own right. The open discussion about what kind of properties would or could make a structure proto-recursive or not will be essential to move the state-of-knowledge past antithetical, jointly exhaustive or mutually exclusive options of what recursion is, how it emerged and evolved.

Second, primate loud calls are functionally analogous to, for instance, bird and whale song. Accordingly, sophisticated signalling in far-related species does not undermine the value of great apes as living models for the study of language and its underpinning neuro-motoric operations, as often claimed (Berwick and Chomsky, 2019; Johan J. Bolhuis et al., 2018; Bolhuis and Wynne, 2009; Lattenkamp and Vernes, 2018; Suzuki et al., 2018; Vernes et al., 2021) based

on the false assumption that great ape vocal research has been at a standstill for the last half-century (cf. Belyk and Brown, 2017; Bianchi et al., 2016; Crockford et al., 2004; Hopkins et al., 2007; Lameira, 2017; Lameira et al., 2016, 2015, 2013; Lameira and Shumaker, 2019; Pereira et al., 2020; Russell et al., 2013; Staes et al., 2017; Taglialatela et al., 2012; Watson et al., 2015; Wich et al., 2009, 2012). Our findings invite results across lineages to be tested with primates in order to explicitly assess their evolutionary relevance within the human clade. This will help avoid prematurely interpreting absence of evidence for evidence of absence in primates.

Third, our findings suggest that, despite criticism, recursive perceptual capacities identified in primates (Watson et al., 2020) may be factual and likely evolved to subserve vocal signals in the species' natural repertoire (though potentially yet unrecognized). This invites for a renewed interest and re-analysis of primate signalling behaviour in the wild (Gabrić, 2021). However, findings also show that it may be too hasty to discuss whether perceptual capacities in primates or birds are equivalent to those engaged in syntax (Watson et al., 2020) or phonology (Rawski et al., 2021). Such classifications may be putting the proverbial cart before the horse; they are based on untested assumptions (e.g., that syntax and phonology evolved as separate "modules", that one attained modern form before the other, et cetera) that may not have applied to proto-recursive ancestors (Kershenbaum et al., 2014; Miyagawa, 2021).

Forth, given that isochrony universally governs music and that recursion is a feature of music, findings suggest a possible evolutionary link between great ape loud calls and vocal music. Loud calling is an archetypal trait in primates (Wich and Nunn, 2002) but is absent in modern humans. Our findings suggest this may not be coincidental. Great ape loud calling may have preceded and subsequently transmuted into modern recursive vocal structures in humans. Given their conspicuousness, loud calls represent one of the most studied aspects of primate vocal behaviour (Wich and Nunn, 2002), but their rhythmic patterns have seldom been characterized with precision (Clink et al., 2020; De Gregorio et al., 2021; Gamba et al., 2016). Besides our analyses, there are remarkably few confirmed cases of isochrony in great apes (but see Raimondi et al., 2023), but the behaviours that have been rhythmically measured with accuracy have been implicated in the evolution of percussion (Fuhrmann et al., 2015) and musical expression (Dufour et al., 2015; Hattori and Tomonaga, 2020), such as social entrainment in chimpanzees in connection with the origin of dance (Lameira et al., 2019) [a capacity once also assumed to be neurologically impossible in great apes (Fitch, 2017b; Patel, 2014)]. This opens the possibility that recursive vocal production and supporting neural procedures were first and foremost a feature of proto-musical expression in human ancestors, later recruited and "re-engineered" for the generation of linguistic combinatorics.

Future studies outlining the distribution of isochrony across primate (vocal) behaviour offer promising new paths to empirically assess the evolution of recursive signal structures in music and language and will help move the needle forward on one the most tantalizing riddles in the evolution of language and cognition. These crucial data and insights will materialise if, as stewards of our planetary co-habitants, humankind secures the survival of these species and the preservation of their natural wild habitat (Estrada et al., 2022, 2017; Laurance, 2013; Laurance et al., 2012).

Methods and Materials

Study site

We conducted our research at the Tuanan Research Station (2°09'S; 114°26'E), Central Kalimantan, Indonesia. Long calls were opportunistically recorded from identified flanged males (*Pongo pygmaeus wurmbii*) using a Marantz Analogue Recorder PMD222 in combination with a Sennheiser Microphone ME 64 or a Sony Digital Recorder TCD-D100 in combination with a Sony Microphone ECM-M907.

Acoustic data extraction

Audio recordings were transferred to a computer with a sampling rate of 44.1 kHz. Seven acoustic measures were extracted directly from the spectrogram window (window type: Hann; 3 dB filter bandwidth: 124 Hz; grid frequency resolution: 2.69 Hz; grid time resolution: 256 samples) by manually drawing a selection encompassing the complete long call (sub)pulse from onset to offset, using Raven interactive sound analysis software (version 1.5, Cornell Lab of Ornithology). These parameters were duration(s), peak frequency (Hz), peak time, peak frequency contour average slope (Hz), peak frequency contour maximum slope (Hz), average entropy (Hz), signal-to-noise ratio (NIST quick method). Please see software's documentation for full description of parameters (https://ravensoundsoftware.com/knowledge-base/pitch-tracking-frequency-contour-measurements/). Acoustic data extraction complemented the classification of long calls elements, both at the pulse and sub-pulse levels, based on close visual and auditory inspection of spectrograms, both based on elements' distinctiveness between each other as well as in relation to the remaining catalogued orangutan call repertoire (Hardus et al., 2009) (see also supplementary audio files). Of these parameters, duration and peak frequency in particular have been shown to be resilient across recording settings(Lameira et al., 2013) and to adequately represent variation in the time and frequency axes (Lameira et al., 2017).

Rhythm data analyses

Inter-onset-intervals (IOI's = t_k) were only calculated from the begin time (s) of each full- and sub-pulse long call elements using Raven interactive sound analysis software, as above explained. t_k was calculated only from subsequent (full/sub) pulse elements of the same type. Ratio values (r_k) were calculated as

$t_k/(t_k+t_{k+1})$. Following the methodology of Roeske et al., 2020 and De Gregorio et al. 2021, to assess the significance of the peaks around isochrony (corresponding to the 0.5 r_k value), we counted the number of r_ks falling inside on-isochrony ranges ($0.440 < r_k < 0.555$) and off-isochrony ranges ($0.400 < r_k < 0.440$ and $0.555 < r_k < 0.600$), symmetrically falling at the right and left sides of 1:1 ratios (0.5 r_k value). We tested the count of on-isochrony r_ks versus the count of off-isochrony r_ks, per pulse type, with a GLMM for negative-binomial family distributions, using *glmmTMB* R library. In particular, we built a *full* model with the count of r_k values as the response variable, the pulse type in interaction with the range the observation fell in (on- or off- isochrony) as predictors. We added an offset weighting the r_k count based on the width of the bin. The individual contribution was set as random factor. We built a *null* model comprising only the offset and the random intercepts. We checked the number of residuals of the full and *null* models, and compared the two models with a likelihood ratio test (Anova with "Chisq" argument). We calculated p-values for each predictor using the R *summary* function and performed pairwise comparisons for each level of the explanatory variables with *emmeans* R package, adjusting all p-values with Bonferroni correction. We checked normality, homogeneity (via function provided by R. Mundry), and number of the residuals. We checked for overdispersion with *performance* R package (Lüdecke et al., 2020). Graphic visualization was prepared using R (Team, 2013) packages *ggplot2* (Wickham, 2009) and *ggridges* (Wilke, 2022). Data reshape and organization were managed with *dplyr* and *tidyr* R packages.

Acoustic data analyses

Permutated discriminant function analysis with cross classification was performed using R and a function provided by Roger Mundry (Mundry and Sommer, 2007). The script was: pdfa.res=pDFA.crossed (test.fac="Sub-pulse-type", contr.fac="Individual.ID", variables=c("Delta.Time", "Peak.Freq", "Peak.Time", "PFC.Avg.Slope", "PFC.Max.Slope", "Avg.Entropy", "SNR.NIST.Quick"), n.to.sel=NULL, n.sel=100, n.perm=1000, pdfa.data=xdata). These analyses assured that long call elements, at the pulse and sub-pulse level, indeed represented biologically distinct categories.

References

Belyk M, Brown S. 2017. The origins of the vocal brain in humans. *Neurosci Biobehav Rev* **77**:177–193. doi:10.1016/j.neubiorev.2017.03.014

Bennett R. 2018. Recursive prosodic words in Kaqchikel (Mayan). *Glossa: a journal of general linguistics* **3**:67. doi:10.5334/gjgl.550

Berwick RC, Chomsky N. 2019. All or nothing: No half-Merge and the evolution of syntax. *PLoS Biol* **17**:e3000539. doi:10.1371/journal.pbio.3000539

Bianchi S, Reyes LD, Hopkins WD, Taglialatela JP, Sherwood CC. 2016. Neocortical grey matter distribution underlying voluntary, flexible vocalizations in chimpanzees. *Scientific reports* **6**:34733. doi:10.1038/srep34733

Bolhuis Johan J, Beckers GJ, Huybregts MA, Berwick RC, Everaert MB. 2018. Meaningful syntactic structure in songbird vocalizations? *PLoS biology* **16**:e2005157. doi:10.1371/journal.pbio.2005157

Bolhuis Johan J., Beckers GJL, Huybregts MAC, Berwick RC, Everaert MBH. 2018. The slings and arrows of comparative linguistics. *PLoS Biol* **16**:e3000019. doi:10.1371/journal.pbio.3000019

Bolhuis JJ, Tattersall I, Chomsky N, Berwick RC. 2015. Language: UG or Not to Be, That Is the Question. *PLoS Biol* **13**:e1002063. doi:10.1371/journal.pbio.1002063

Bolhuis JJ, Wynne CD. 2009. Can evolution explain how minds work? *Nature* **458**:832–833. doi:10.1038/458832a

Bowling DL, Fitch TW. 2015. Do Animal Communication Systems Have Phonemes? *Trends in Cognitive Sciences* **19**:555–557. doi:10.1016/j.tics.2015.08.011

Chomsky N. 2010. Some simple evo devo theses: how true might they be for language? In: Larson RK, Deprez V, Yamakido H, editors. The Evolution of Human Language. Cambridge: Cambridge University Press. pp. 45–62. doi:10.1017/CBO9780511817755.003

Clink DJ, Tasirin JS, Klinck H. 2020. Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Current Zoology* **66**:173–186. doi:10.1093/cz/zoz035

Corballis MC, Corballis MC. 2014. The Recursive Mind: The Origins of Human Language, Thought, and Civilization - Updated Edition. Princeton University Press. doi:10.1515/9781400851492

Crockford C, Herbinger I, Vigilant L, Boesch C. 2004. Wild Chimpanzees Produce Group-Specific Calls: a Case for Vocal Learning? *Ethology* **110**:221–243. doi:10.1111/j.1439-0310.2004.00968.x

de Boer B, Thompson B, Ravignani A, Boeckx C. 2020. Evolutionary Dynamics Do Not Motivate a Single-Mutant Theory of Human Language. *Sci Rep* **10**:451. doi:10.1038/s41598-019-57235-8

De Gregorio C, Valente D, Raimondi T, Torti V, Miaretsoa L, Friard O, Giacoma C, Ravignani A, Gamba M. 2021. Categorical rhythms in a singing primate. *Current Biology* **31**:R1379–R1380. doi:10.1016/j.cub.2021.09.032

Dufour V, Poulin N, Curé C, Sterck EH. 2015. Chimpanzee drumming: a spontaneous performance with characteristics of human musical drumming. *Scientific reports* **5**:11320. doi:10.1038/srep11320

Elfner E. 2015. Recursion in prosodic phrasing: evidence from Connemara Irish. *Nat Lang Linguist Theory* **33**:1169–1208. doi:10.1007/s11049-014-9281-5

Engesser S, Holub JL, O'Neill LG, Russell AF, Townsend SW. 2019. Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *PNAS* 201819513. doi:10.1073/pnas.1819513116

Engesser S, Ridley AR, Townsend SW. 2016. Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences of the United States of America* 201600970. doi:10.1073/pnas.1600970113

Estrada A, Garber PA, Gouveia S, Fernández-Llamazares Á, Ascensão F, Fuentes A, Garnett ST, Shaffer C, Bicca-Marques J, Fa JE, Hockings K, Shanee S, Johnson S, Shepard GH, Shanee N, Golden CD, Cárdenas-Navarrete A, Levey DR, Boonratana R, Dobrovolski R, Chaudhary A, Ratsimbazafy J, Supriatna J, Kone I, Volampeno S. 2022. Global importance of Indigenous Peoples, their lands, and knowledge systems for saving the world's primates from extinction. *Sci Adv* **8**:eabn2927. doi:10.1126/sciadv.abn2927

Estrada A, Garber PA, Rylands AB, Roos C, Eduardo F-D, Fiore A, Nekaris A-IK, Nijman V, Heymann EW, Lambert JE, Rovero F, Barelli C, Setchell JM, Gillespie TR, Mittermeier RA, Arregoitia L, de Guinea M, Gouveia S, Dobrovolski R, Shanee S, Shanee N, Boyle SA, Fuentes A, C M Katherine, Amato KR, Meyer AL, Wich S, Sussman RW, Pan R, Kone I, Li B. 2017. Impending extinction crisis of the world's primates: Why primates matter e1600946. doi:10.1126/sciadv.1600946

Ferrigno S, Cheyette SJ, Piantadosi ST, Cantlon JF. 2020. Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Sci Adv* **6**:eaaz1002. doi:10.1126/sciadv.aaz1002

Fitch TW. 2017a. Dendrophilia and the Evolution of Syntax In: Boë L-J, Joël F, Pascal P, Jean-Luc S, editors. Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates. Berlin: Peter Lang. pp. 305–328.

Fitch TW. 2017b. Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review* **24**:1–31. doi:10.3758/s13423-017-1236-5

Fuhrmann D, Ravignani A, Sarah M-P, Whiten A. 2015. Synchrony and motor mimicking in chimpanzee observational learning. *Sci Rep-uk* **4**:srep05283. doi:10.1038/srep05283

Gabrić P. 2021. Overlooked evidence for semantic compositionality and signal reduction in wild chimpanzees (Pan troglodytes). *Anim Cogn*. doi:10.1007/s10071-021-01584-3

Gamba M, Torti V, Estienne V, Randrianarison R, Valente D, Rovara P, Bonadonna G, Friard O, Giacoma C. 2016. The Indris Have Got Rhythm! Timing and Pitch Variation Of A Primate Song Examined Between Sexes And Age Classes. *Frontiers in neuroscience* **10**:249. doi:10.3389/fnins.2016.00249

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. 2006. Recursive syntactic pattern learning by songbirds. *Nature* **440**:1204–1207. doi:10.1038/nature04675

Hardus ME, Lameira AR, Singleton I, C M-B Helen, Knott CD, Ancrenaz M, Utami S, Wich Serge. 2009. A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation In: Wich S, Setia MT, Utami SS, Schaik C, editors. Orangutans. New York: Oxford University Press. pp. 49–60.

Hattori Y, Tomonaga M. 2020. Rhythmic swaying induced by sound in chimpanzees (*Pan troglodytes*). *Proc Natl Acad Sci USA* **117**:936–942. doi:10.1073/pnas.1910318116

Hauser MD, Chomsky N, Fitch TW. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science (New York, NY)* **298**:1569–1579. doi:10.1126/science.298.5598.1569

Hockett CF. 1960. The origin of speech. *Scientific American* **203**:89–96.

Hopkins WD, Taglialatela JP, Leavens DA. 2007. Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour* **73**:281–286.

Idsardi WJ, Gallego ÁJ, Martin R. 2018. Why Is Phonology Different? No Recursion In: Gallego ÁJ, Martin R, editors. Language, Syntax, and the Natural Sciences. Cambridge University Press. pp. 212–223.

Jackendoff R. 2009. Parallels and Nonparallels between Language and Music. *Music Perception* **26**:195–204. doi:10.1525/mp.2009.26.3.195

Jackendoff R, Pinker S. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* **97**:211–225. doi:10.1016/j.cognition.2005.04.006

Jiang X, Long T, Cao W, Li J, Dehaene S, Wang L. 2018. Production of Supra-regular Spatial Sequences by Macaque Monkeys. *Current Biology* **0**:1851-1859.e4. doi:10.1016/j.cub.2018.04.047

Kabak Barış, Revithiadou A. 2009. An interface approach to prosodic word recursion In: Grijzenhout J, Kabak Baris, editors. Phonological Domains, Interface Explorations. Berlin, New York: Mouton de Gruyter. pp. 105–134. doi:10.1515/9783110219234.2.105

Kershenbaum A, Bowles AE, Freeberg TM, Jin DZ, Lameira AR, Bohn K. 2014. Animal vocal sequences: not the Markov chains we thought they were. *Proceedings Biological sciences / The Royal Society* **281**:20141370. doi:10.1098/rspb.2014.1370

Koelsch S, Rohrmeier M, Torrecuso R, Jentschke S. 2013. Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America* **110**:15443–15448. doi:10.1073/pnas.1300272110

Lameira AR. 2017. Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neuroscience & Biobehavioral Reviews* **83**:429–439. doi:10.1016/j.neubiorev.2017.09.021

Lameira AR, Alexandre A, Gamba M, Nowak MG, Vicente R, Wich S. 2021. Orangutan information broadcast via consonant-like and vowel-like calls breaches mathematical models of linguistic evolution. *Biol Lett* **17**:20210302. doi:10.1098/rsbl.2021.0302

Lameira AR, Eerola T, Ravignani A. 2019. Coupled whole-body rhythmic entrainment between two chimpanzees. *Sci Rep* **9**:18914. doi:10.1038/s41598-019-55360-y

Lameira AR, Hardus ME, Bartlett AM, Shumaker RW, Wich SA, Menken SB. 2015. Speech-like rhythm in a voiced and voiceless orangutan call. *PloS one* **10**:e116136. doi:10.1371/journal.pone.0116136

Lameira AR, Hardus ME, Kowalsky B, de Vries H, Spruijt BM, Sterck E, Shumaker RW, Wich S. 2013. Orangutan (Pongo spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *Journal of the Acoustical Society of America* **134**:1–11. doi:10.1121/1.4817929

Lameira AR, Hardus ME, Mielke A, Wich SA, Shumaker RW. 2016. Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific reports* **6**:30315. doi:10.1038/srep30315

Lameira AR, Shumaker RW. 2019. Orangutans show active voicing through a membranophone. *Sci Rep* **9**:12289. doi:10.1038/s41598-019-48760-7

Lameira AR, Vicente R, Alexandre A, Gail C-S, Knott C, Wich S, Hardus ME. 2017. Proto-consonants were information-dense via identical bioacoustic tags to proto-vowels. *Nature Human Behaviour* **1**:0044. doi:10.1038/s41562-017-0044

Lameira AR, Wich S. 2008. Orangutan Long Call Degradation and Individuality Over Distance: A Playback Approach. *International Journal of Primatology* **29**:615–625. doi:10.1007/s10764-008-9253-x

Lattenkamp EZ, Vernes SC. 2018. Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Curr Opin Behav Sci.* doi:10.1016/j.cobeha.2018.04.007

Laurance WF. 2013. Does research help to safeguard protected areas? *Trends in Ecology {& Evolution* **28**:261–266. doi:10.1016/j.tree.2013.01.017

Laurance WF, Useche CD, Rendeiro J, Kalka M, Bradshaw CJ, Sloan SP, Laurance SG, Campbell M, Abernethy K, Alvarez P, Victor A-R, Ashton P, Julieta B-M, Blom A, Bobo KS, Cannon CH, Cao M, Carroll R, Chapman C, Coates R, Cords M, Danielsen F, Dijn B, Dinerstein E, Donnelly MA, Edwards D, Edwards F, Farwig N, Fashing P, Forget P-M, Foster M, Gale G, Harris D, Harrison R, Hart J, Karpanty S, Kress JW, Krishnaswamy J, Logsdon W, Lovett J, Magnusson W, Maisels F, Marshall AR, Deedra

M, Mudappa D, Nielsen MR, Pearson R, Pitman N, van der Ploeg J, Plumptre A, Poulsen J, Quesada M, Rainey H, Robinson D, Roetgers C, Rovero F, Scatena F, Schulze C, Sheil D, Struhsaker T, Terborgh J, Thomas D, Timm R, J U-C Nicolas, Vasudevan K, Wright JS, Juan A-G, Arroyo L, Ashton M, Auzel P, Babaasa D, Babweteera F, Baker P, Banki O, Bass M, Inogwabini B-I, Blake S, Brockelman W, Brokaw N, Bruhl CA, Bunyavejchewin S, Chao J-T, Chave J, Chellam R, Clark CJ, Clavijo J, Congdon R, Corlett R, Dattaraja H, Dave C, Davies G, de Beisiegel B, de da Silva R, Fiore A, Diesmos A, Dirzo R, Diane D-S, Eaton M, Emmons L, Estrada A, Ewango C, Fedigan L, Feer F, Fruth B, Willis J, Goodale U, Goodman S, Guix JC, Guthiga P, Haber W, Hamer K, Herbinger I, Hill J, Huang Z, Sun FI, Ickes K, Itoh A, Ivanauskas N, Jackes B, Janovec J, Janzen D, Jiangming M, Jin C, Jones T, Justiniano H, Kalko E, Kasangaki A, Killeen T, King H, Klop E, Knott C, Kone I, Kudavidanage E, da Ribeiro J, Lattke J, Laval R, Lawton R, Leal M, Leighton M, Lentino M, Leonel C, Lindsell J, Lee L-L, Linsenmair EK, Losos E, Lugo A, Lwanga J, Mack AL, Martins M, W M Scott, Roan M, Montag L, Thompson J, Jacob N-N, Nakagawa M, Nepal S, Norconk M, Novotny V, Sean O, Opiang M, Ouboter P, Parker K, Parthasarathy N, Pisciotta K, Prawiradilaga D, Pringle C, Rajathurai S, Reichard U, Reinartz G, Renton K, Reynolds G, Reynolds V, Riley E, Rodel M-O, Rothman J, Round P, Sakai S, Sanaiotti T, Savini T, Schaab G, Seidensticker J, Siaka A, Silman MR, Smith TB, de Almeida S, Sodhi N, Stanford C, Stewart K, Stokes E, Stoner KE, Sukumar R, Surbeck M, Tobler M, Tscharntke T, Turkalo A, Umapathy G, van Weerd M, Rivera J, Venkataraman M, Venn L, Verea C, de Castilho C, Waltert M, Wang B, Watts D, Weber W, West P, Whitacre D, Whitney K, Wilkie D, Williams S, Wright DD, Wright P, Xiankai L, Yonzon P, Zamzani F. 2012. Averting biodiversity collapse in tropical forest protected areas. *Nature* **advance on**. doi:10.1038/nature11318

Liao DA, Brecht KF, Johnston M, Nieder A. 2022. Recursive sequence generation in crows. *Sci Adv* **8**:eabq3356. doi:10.1126/sciadv.abq3356

Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahasi M, Suzuki K, Feher O, Ravbar P, Okanoya K, Tchernichovski O. 2013. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* **498**:104–108. doi:10.1038/nature12173

Mandelbrot BB. 1980. FRACTAL ASPECTS OF THE ITERATION OF z →Λz(1- z) FOR COMPLEX Λ AND z. *Annals of the New York Academy of Sciences* **357**:249–259. doi:10.1111/j.1749-6632.1980.tb29690.x

Martins M. 2012. Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**:2055–2064. doi:10.1098/rstb.2012.0097

Martins MD, Gingras B, Puig-Waldmueller E, Fitch WT. 2017. Cognitive representation of "musical fractals": Processing hierarchy and recursion in the auditory domain. *Cognition* **161**:31–45. doi:10.1016/j.cognition.2017.01.001

Martins PT, Boeckx C. 2019. Language evolution and complexity considerations: The no half-Merge fallacy. *PLoS Biol* **17**:e3000389. doi:10.1371/journal.pbio.3000389

Miyagawa S. 2021. Revisiting Fitch and Hauser's Observation That Tamarin Monkeys Can Learn Combinations Based on Finite-State Grammar. *Front Psychol* **12**:772291. doi:10.3389/fpsyg.2021.772291

Mundry R, Sommer C. 2007. Discriminant function analysis with nonindependent data: consequences and an alternative. *Animal Behaviour* **74**:965–976. doi:10.1016/j.anbehav.2006.12.028

Nasukawa K, editor. 2020. Morpheme-internal recursion in phonology, Studies in generative grammar. Berlin ; Boston: De Gruyter Mouton.

Nasukawa K. 2015. Recursion in the lexical structure of morphemes In: Oostendorp M van, Riemsdijk H van, editors. Representing Structure in Phonology and Syntax. DE GRUYTER. pp. 211–238. doi:10.1515/9781501502224-009

Patel AD. 2014. The evolutionary biology of musical rhythm: was Darwin wrong? *PLoS biology* **12**:e1001821. doi:10.1371/journal.pbio.1001821

Pereira AS, Kavanagh E, Hobaiter C, Slocombe KE, Lameira AR. 2020. Chimpanzee lip-smacks confirm primate continuity for speech-rhythm evolution. *Biol Lett* **16**:20200232. doi:10.1098/rsbl.2020.0232

Pouw W, Paxton A, Harrison SJ, Dixon JA. 2020. Acoustic information about upper limb movement in voicing. *Proc Natl Acad Sci USA* 202004163. doi:10.1073/pnas.2004163117

Quam R, Martínez I, Rosa M, Bonmatí A, Lorenzo C, de Ruiter DJ, Jacopo M-C, Valverde M, Jarabo P, Menter CG, Thackeray FJ, Arsuaga J. 2015. Early hominin auditory capacities {\textbar} Science Advances. *Sci Adv* **1**:e1500355. doi:10.1126/sciadv.1500355

Raimondi T, Di Panfilo G, Pasquali M, Zarantonello M, Favaro L, Savini T, Gamba M, Ravignani A. 2023. Isochrony and rhythmic interaction in ape duetting. *Proc R Soc B* **290**:20222244. doi:10.1098/rspb.2022.2244

Rawski J, Idsardi W, Heinz J. 2021. Comment on "Nonadjacent dependency processing in monkeys, apes, and humans." *Sci Adv* **7**:eabg0455. doi:10.1126/sciadv.abg0455

Roeske TC, Tchernichovski O, Poeppel D, Jacoby N. 2020. Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology* **30**:3544-3555.e6. doi:10.1016/j.cub.2020.06.072

Russell JL, Joseph M, Hopkins WD, Taglialatela JP. 2013. Vocal learning of a communicative signal in captive chimpanzees, Pan troglodytes. *Brain and Language* **127**:520–525. doi:10.1016/j.bandl.2013.09.009

Sainburg T, Theilman B, Thielk M, Gentner TQ. 2019. Parallels in the sequential organization of birdsong and human speech. *Nat Commun* **10**:1–11. doi:10.1038/s41467-019-11605-y

Sharma N, Chimalakonda S. 2018. Learning Recursion from Music and Music from Recursion2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT). Presented at the 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT). Mumbai: IEEE. pp. 257–261. doi:10.1109/ICALT.2018.00066

Spillmann B, Dunkel LP, van Noordwijk MA, Amda RN, Lameira AR, Wich SA, van Schaik CP. 2010. Acoustic properties of long calls given by flanged male orang-utans (Pongo pygmaeus wurmbii) reflect both individual identity and context. *Ethology* **116**:385–395.

Staes N, Sherwood CC, Wright K, de Manuel M, Guevara EE, Tomas M-B, Krützen M, Massiah M, Hopkins WD, Ely JJ, Bradley BJ. 2017. FOXP2 variation in great ape populations offers insight into the evolution of communication skills. *Scientific reports* **7**:258. doi:10.1038/s41598-017-16844-x

Suzuki T, Wheatcroft D, communications G-M. 2016. Experimental evidence for compositional syntax in bird calls. *Nature communications*.

Suzuki TN, Wheatcroft D, Griesser M. 2018. Call combinations in birds and the evolution of compositional syntax. *PLoS Biol* **16**:e2006532. doi:10.1371/journal.pbio.2006532

Suzuki TN, Wheatcroft D, Griesser M. 2017. Wild Birds Use an Ordering Rule to Decode Novel Call Sequences. *Current Biology* **27**:2331-2336.e3. doi:10.1016/j.cub.2017.06.031

Taglialatela JP, Reamer L, Schapiro SJ, Hopkins WD. 2012. Social learning of a communicative signal in captive chimpanzees. *Biology letters* **8**:498–501. doi:10.1098/rsbl.2012.0113

Team R. 2013. R: A language and environment for statistical computing.

Vernes SC, Janik VM, Fitch WT, Slater PJB. 2021. Vocal learning in animals and humans. *Phil Trans R Soc B* **376**:20200234. doi:10.1098/rstb.2020.0234

Vogel I. 2012. Recursion in phonology? In: Botma B, Noske R, editors. Phonological Explorations. Berlin, Boston: DE GRUYTER. doi:10.1515/9783110295177.41

von Humboldt W. 1836. Über die Verschiedenheit des Menschlichen Sprachbaues und ihren Einfluss auf die geristige Entwickelung des Menschengeschlechts. Berlin: Königlichen Akademie der Wissenschaften.

Wang L, Uhrig L, Jarraya B, Dehaene S. 2015. Representation of Numerical and Sequential Patterns in Macaque and Human Brains. *Current Biology* **25**:1966–1974. doi:10.1016/j.cub.2015.06.035

Watson SK, Burkart JM, Schapiro SJ, Lambeth SP, Mueller JL, Townsend SW. 2020. Nonadjacent dependency processing in monkeys, apes, and humans. *Sci Adv* **6**:eabb0725. doi:10.1126/sciadv.abb0725

Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, West V, Slocombe KE. 2015. Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology* **25**:495–499. doi:10.1016/j.cub.2014.12.032

Wich S, Swartz K, Hardus M, Lameira A, Stromberg E, Shumaker R. 2009. A case of spontaneous acquisition of a human sound by an orangutan. *Primates* **50**:56–64. doi:10.1007/s10329-008-0117-y

Wich SA, KRU T MICHAEL, Lameira AR, ALEXANDER N, NATASHA A, Bastian ML, Meulman E, C M-B Helen, Atmoko SS, Pamungkas J, Dyah P-F, Hardus ME, van Noordwijk M, van Schaik CP. 2012. Call cultures in orang-utans? *PloS one* **7**:e36180. doi:10.1371/journal.pone.0036180

Wich SA, Nunn CL. 2002. Do male "long-distance calls" function in mate defense? A comparative study of long-distance calls in primates. *Behavioral Ecology and Sociobiology* **52**:474–484. doi:10.1007/s00265-002-0541-8

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

Wilke C. 2022. ggridges.

Acknowledgements

We thank the Indonesian Ministry of Research and Technology, the Indonesian Ministry of Environment and Forestry, the Indonesian Ministry of Home Affairs, the Directorate General of Natural Resources and Ecosystem Conservation and the former Directorate General of Forest Protection and Nature Conservation for authorization to carry out research in Indonesia; the Universitas National for supporting the project and acting as sponsors and counter-partners; the Bornean Orangutan Survival Foundation and the MAWAS Programme in Palangkaraya for their support and permission to stay and work in the MAWAS Reserve. A.R.L. was supported by the UK Research & Innovation, Future Leaders Fellowship grant agreement number MR/T04229X/1.

Author contributions

A.R.L. conceived and designed the study. A.R.L. and M.E.H. collected data. A.R.L., A.R., T.R. and M.G. analysed data. A.R.L., M.E.H., A.R., T.R. and M.G. wrote the paper.

Competing interests

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to Adriano R. Lameira.

Figure 1 – Organization and rhythmic features of orangutans' long calls.

(A) On top, the spectrogram of a *Full Pulse* and its organization in Sub-Pulses (e.g., *Grumble sub-pulses*). Below are the spectrograms of the three other sub-element types: *Sub-pulse transitory* elements, *Pulse bodies* and *Bubble sub-pulses*. Bars on the top of each spectrogram schematically quantify durations of inter-onset intervals (t_k): dark green denotes the higher-level of organization (*Full pulse*). Orange (in the inset) and light green (bottom right) denote the lower-level organization (sub-pulse element types).

(B) Probability density function showing the distributions of the inter-onset-intervals (t_k) for each of the long call element types.

(C) The distributions on the left show rhythm ratios (r_k) per element type as calculated on 12 flanged males for a total of 1915 Full-pulses and 5309 sub-pulses. Solid sections of the curves indicate on-isochrony r_k values; striped sections indicate off-isochrony r_k values. A solid white line indicates the 0.5 r_k value corresponding to isochrony. White dotted lines denote the on-isochrony peak value extracted from the probability density function. On the right, a bar plot per each element type shows the percentage of observations (r_k) falling into the on-isochrony boundaries (solid bars) or on off-isochrony boundaries (striped bars). The number of on-isochrony r_k is significantly larger (GLMM, *Full* vs *Null*: Chisq=2717.543, $p<0.001$) than the number of off-isochrony r_k for all long call element types (*Full pulse*: t-ratio=-25.164, $p<0.001$; *Bubble sub-pulse*: t-ratio= -30.694, $p<0.001$; *Grumble sub-pulse*: t-ratio=-14.526, $p<0.001$; *Sub-pulse transitory element*: t-ratio=-3.148, $p<0.001$). *Pulse body* showed no r_k values falling within the on-off-isochrony boundaries.

(D) Distribution of a variable calculated as the ratio between the t_k of a sub-pulse and the t_k of the corresponding higher level of organization, the *Full Pulse*. We report the peak value of the curve (0.046) and tested the significance of the extent of the central quartiles, which was significantly smaller than peripheral quartiles (Wilcoxon signed-rank test: W=2272, $p<0.001$).

Supplementary Materials

Count of obs ~ Pulse level *Type + (1|ID Contribution)

Pulse level: BSP, FP, GSP, SBTE

Type: ON or OFF isochrony

Model	Df	AIC	BIC	logLik	deviance	Chisq	ChiDf	Pr(>Chisq)
null	4	2398.896	2414.092	-1195.448	2390.896	-	-	-
full	11	2114.608	2156.398	-1046.304	2092.608	298.288	7	p<0.001

Random effects

Conditional model

Groups	Name	Variance	Std.Dev.
ID contribution	(Intercept)	0.2705	0.5201

Number of obs: 330. groups: ID contribution, 65

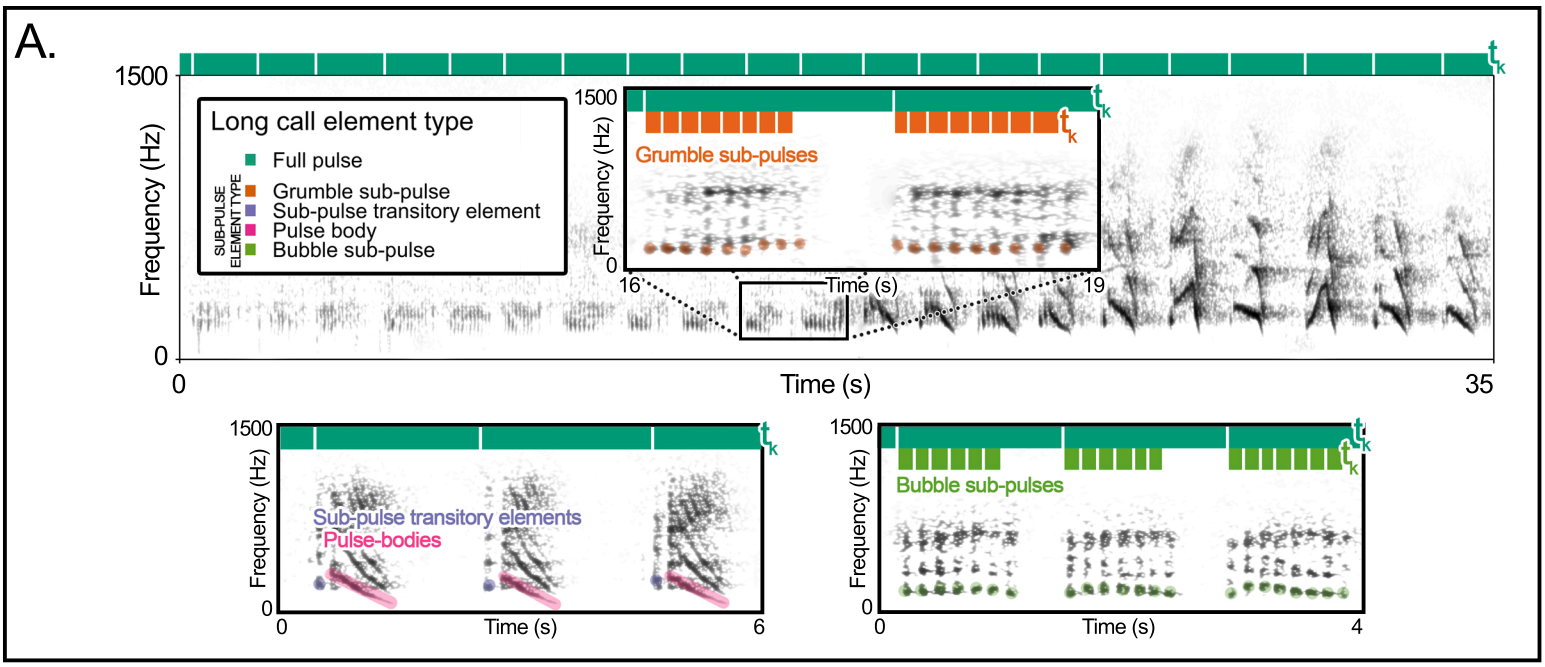
Conditional model:

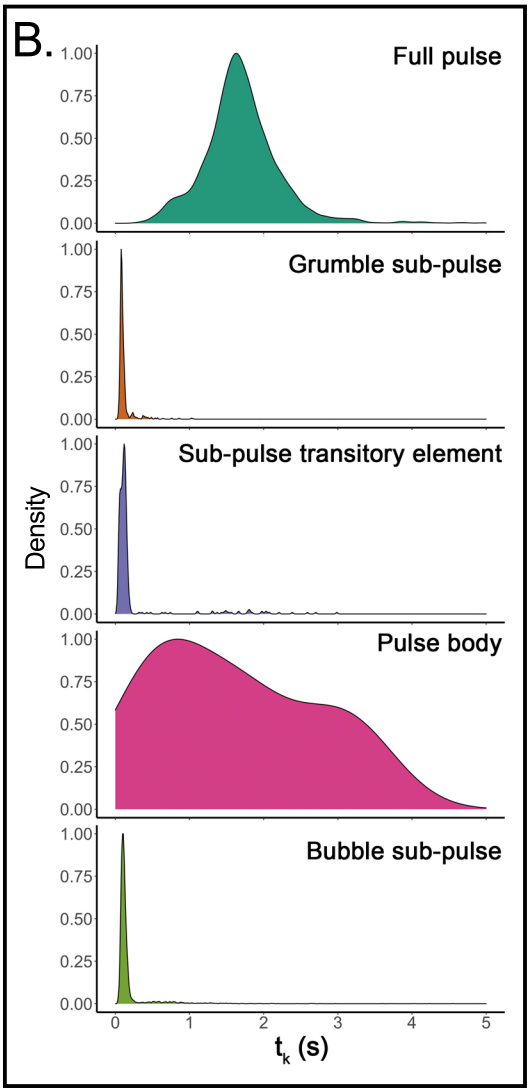
	Estimate	Std Error	zvalue	Pr(>|z|)	
(Intercept)					
Type (ON) [a]	1.2959	0.1445	8.97	<0.0001	***
Pulse level (FP)[a]	-1.3532	0.1670	-8.10	<0.001	***
Pulse level (GSP) [a]	-0.8826	0.2449	-3.60	<0.001	***
Pulse level (SBTE) [a]	-2.3268	0.2966	-7.85	<0.001	***
Pulse level (FP) : type (ON) [a]	1.0314	0.2150	4.80	<0.001	***
Pulse level (GSP) : type (ON) [a]	0.1615	0.3241	0.50	0.618283	
Pulse level (SBTE) : type (ON)[a]	-0.1903	0.3626	-0.52	0.599660	

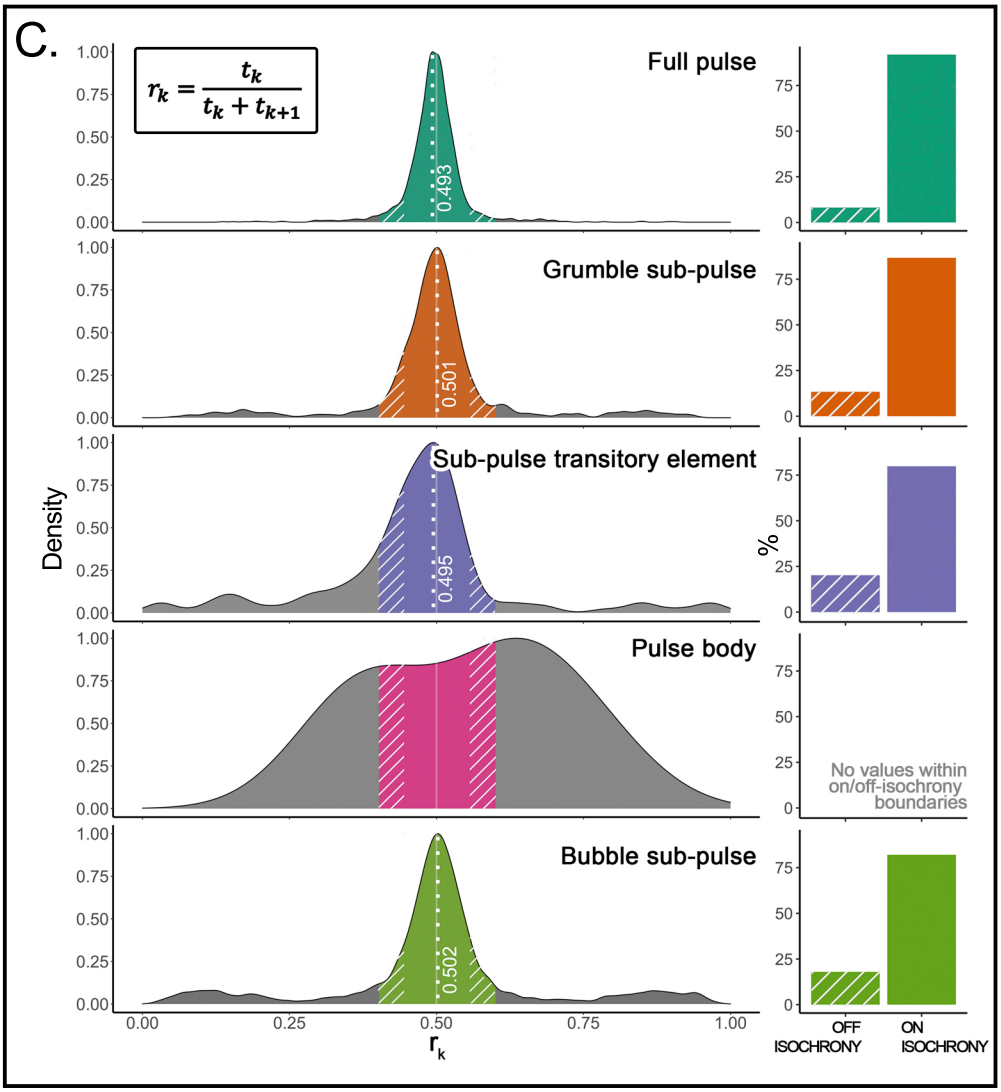
[a]*Pulse level (BSP) and Type (OFF) being the reference categories*

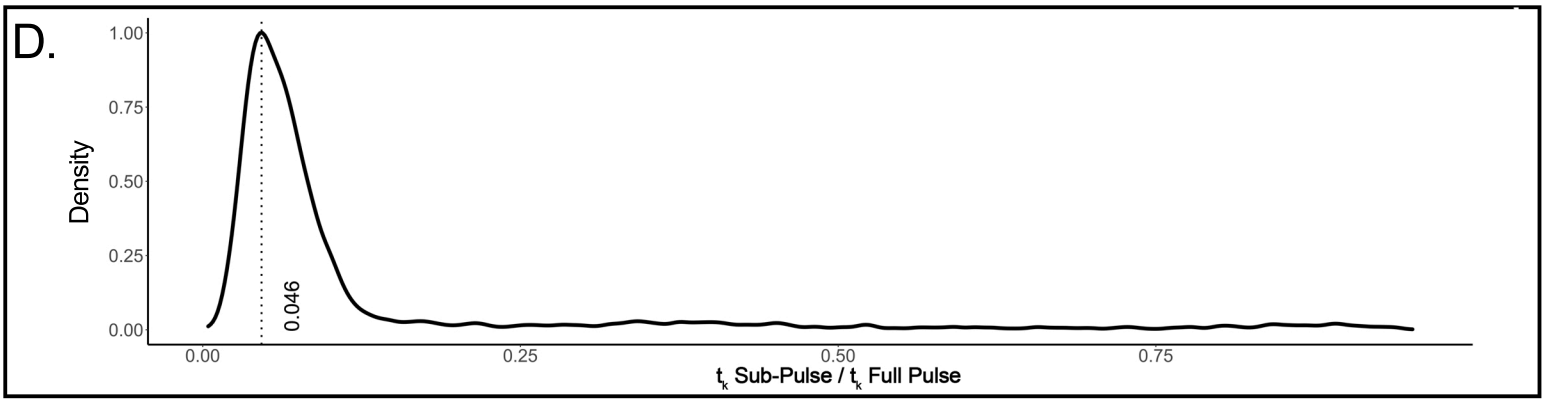
***Post-hoc* test**

Contrast	Estimate	SE	df	t.ratio	p.value
FP: OFF - ON	-2.57	0.161	319	-15.957	<.0001
BSP: OFF - ON	-1.54	0.144	319	-10.640	<.0001
GSP: OFF - ON	-1.70	0.286	319	-5.940	<.0001
SBTE: OFF - ON	-1.35	0.333	319	-4.048	0.0001


A.
Long call element type
SUB-PULSE ELEMENT TYPE
Full pulse
Grumble sub-pulse
Sub-pulse transitory element
Pulse body
Bubble sub-pulse
Frequency (Hz)
Time (s)
Grumble sub-pulses
Sub-pulse transitory elements
Pulse-bodies
Bubble sub-pulses


B.
Full pulse
Grumble sub-pulse
Sub-pulse transitory element
Pulse body
Bubble sub-pulse
Density
t_k (s)


C.
$r_k = \frac{t_k}{t_k + t_{k+1}}$
Full pulse
Grumble sub-pulse
Sub-pulse transitory element
Pulse body
Bubble sub-pulse
0.493
0.501
0.495
0.502
Density
%
No values within on/off-isochrony boundaries
r_k
OFF ISOCHRONY
ON ISOCHRONY


D.
Density
0.046
t_k Sub-Pulse / t_k Full Pulse