RESEARCH ARTICLE

An updated vocal repertoire of wild adult bonobos (*Pan paniscus*)

Franziska Wegdell[1,2*], **Isaac Schamberg**[1,2], **Mélissa Berthet**[1,2], **Yannik Rothacher**[3], **Volker Dellwo**[4], **Martin Surbeck**[5,6☯], **Simon W. Townsend**[1,2☯]

1 Department of Evolutionary Anthropology, University of Zurich, Zurich, Switzerland, **2** Institute for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland, **3** Department of Psychology, University of Zurich, Zurich, Switzerland, **4** Department of Computational Linguistics, University of Zurich, Zurich, Switzerland, **5** Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America, **6** Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

☯ These authors contributed equally to this work.
* franziska.wegdell@outlook.com

Check for updates

OPEN ACCESS

Citation: Wegdell F, Schamberg I, Berthet M, Rothacher Y, Dellwo V, Surbeck M, et al. (2025) An updated vocal repertoire of wild adult bonobos (*Pan paniscus*). PLoS One 20(9): e0330250. https://doi.org/10.1371/journal.pone.0330250

Editor: Sidarta Ribeiro, Federal University of Rio Grande do Norte: Universidade Federal do Rio Grande do Norte, BRAZIL

Received: March 1, 2024

Accepted: July 28, 2025

Published: September 10, 2025

Copyright: © 2025 Wegdell et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All data and code files are available from the OSF database: https://osf.io/wes3f/?view_only=53206fc-8d6a047cc9be409da4b4a816a.

Funding: FW, SWT and VD were supported by the NCCR Evolving Language, SwCSS NSF

Abstract

Research over the last 20 years has shed important light on the vocal behaviour of our closest living relatives, bonobos and chimpanzees, but mostly relies on qualitative vocal repertoires, for which quantitative validations are absent. Such data are critical for a holistic understanding of a species` communication system and unpacking how these systems compare more broadly with other primate and non-primate species. Here we make key progress by providing the first quantitative validation of a *Pan* vocal repertoire, specifically for wild bonobos. Using data comprising over 1500 calls from 53 adult individuals collected over 33 months, we employ machine-learning-based random forest analyses and describe 11 acoustically distinguishable call types. We discuss issues associated with resolving vocal repertoires from wild data in great apes and highlight potential future approaches to further capture the complexity and gradedness of the bonobo vocal system.

Introduction

Over the last 20 years, considerable research attention has focused on the vocal behaviour of our closest-living relatives, non-human great apes, not least given the insights such findings can provide into the evolution of our own communication system: language. Unsurprisingly, a significant proportion of work to date has focused on our closest cousins, the *Pan* lineage (chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*)). Research in chimpanzees has demonstrated that they use calls flexibly across contexts [1]; possess a degree of plasticity within call production [2,3]; deploy calls voluntarily [4–9]; produce extensive repertoires of call combinations [10–14]; extract meaning from such combinations [15] and the call order can differ between populations [16]. Similar findings have been demonstrated in bonobos;

Agreement Nr.51NF40_180888. SWT was supported by the Swiss National Science Foundation (Grant: PP00P3_163850 & PP00P3_198912), MB was supported by PP00P3_198912 to SWT. IS was supported by the Swiss National Science Foundation (Grant: 315130_192620, to SWT). IS was additionally supported by the Leakey Foundation and MS was supported by the startup budget granted from Harvard University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

individuals also deploy calls flexibly across contexts [17], deploy calls voluntarily [5] use call combinations extensively [18–20] and can derive meaning from them [21,22]. In addition, certain calls demonstrate functional flexibility (i.e., where calls can be used across contexts characterized by different emotional valence) [23] and populations differ in call and call combination usage [24,25]. Despite considerable progress in decomposing the complexity of *Pan* vocal systems, in both wild chimpanzees and bonobos, only qualitative descriptions of their vocal repertoires exist. Potential reasons for this include, but are not limited to, the fact that *Pan* vocal repertoires are graded [1], where calls overlap in a continuous acoustic space and quantitative approaches to resolving a call repertoire necessarily require a large and robust sample size. Here, we make key progress on this issue by validating descriptive repertoires and providing the first quantitative *Pan* vocal repertoire using calls specifically from wild bonobos.

Previous attempts to describe the bonobo vocal repertoire exist. First, de Waal [26] took a qualitative approach (with calls being classed into call type categories by human ear or by visually assessing spectrograms) recording 10 captive, mainly immature individuals and described 17 call types. Hopkins & Savage-Rumbaugh [27] described 10 call types for group-living captive bonobos and 14 for a human-reared bonobo. Almost a decade later, Bermejo & Omedes [28] followed this up and published the only vocal repertoire on wild bonobos to date, which again qualitatively, described 15 call types. More recently, Keenan and colleagues [29] and Arnaud and colleagues [30] have revisited the vocal repertoire of captive bonobos and, although quantitative methods were employed, only a subset of calls from the repertoire (namely *high hoots*, *barks*, *soft barks*, *peep-yelps* and *peeps*) were considered.

Whilst qualitative repertoires serve as an important entry point facilitating the investigation into non-human animal vocal behaviour, they are unavoidably subjective and susceptible to bias since they rely purely on human visual and auditory discrimination. One consequence of this subjective approach is that various researchers seem to carve up the repertoire of the same species differently. Bonobos are no exception and hence much inconsistency between studies regarding the naming and definition of call types exists. Bermejo and Omedes [28] for example used the term *barks* for calls that other studies [20,26,31] termed *high hoots*; and did not distinguish *contest hoots* (described in [26] and [17]) from other types of hoots. As another example, *soft barks* were described by Bermejo and Omedes [28] and Keenan [29], but Clay and Zuberbühler [21] used the label *food barks,* and de Waal [26] did not describe this call type at all. Quantitative vocal repertoires, whereby calls are classified based on their acoustic features, represent one way to pressure test and subsequently validate existing qualitative repertoires. Such quantitative approaches further allow specific questions that were previously intractable to be addressed. For example, quantitative repertoires can be compared more reliably across individuals, groups, populations and species. Additionally, quantitative repertoires can facilitate and optimize conservation-driven questions such as passive acoustic monitoring-based population surveys where knowledge regarding the vocal repertoire of a species is needed [32] and which are more efficient than camera traps in detecting primates [33].

Quantitative approaches to capture vocal repertoires range from traditional acoustic analysis and discrimination-based statistical analyses [34–36], to more state-of-the-art machine learning-based methods where supervised or unsupervised algorithms cluster calls into defined or undefined categories, respectively [29,37,38]. Here we attempt to quantitatively assess and validate the hitherto described call types of the vocal repertoire of wild adult bonobos specifically using machine learning-based supervised random forest classifying algorithms.

Materials and methods

Ethics statement

Ethical permission to conduct this non-invasive study was granted by the Institut Congolais pour la Conservations de la Nature and the Ministry of Research and Technology of the Democratic Republic of the Congo. This study is in line with the ethical guidelines of the former Department of Primatology at the Max-Planck-Institute for Evolutionary Anthropology and the guidelines of the American Society of Primatologists for the ethical treatment of non-human primates. We also adhered to the best practice guidelines for health monitoring and disease control in great ape populations [39]: researchers underwent quarantine, wore masks during data collection and maintained a 7m distance to the bonobos. Access to the Kokolopori Bonobo Reserve was granted by the villages of Bolamba, Yete, Yomboli, and Yasalakose. Additional information regarding the ethical, cultural, and scientific considerations specific to inclusivity in global research is included in the Supporting Information (S7 Questionnaire).

Data collection

We recorded vocalisations using *ad libitum* and focal sampling methods [40] for 33 months (2107 recording hours) over a period of 10 years (2011–2022) from habituated wild adult bonobos. Subjects were recorded at two field sites in the Democratic Republic of Congo, the Kokolopori Bonobo Reserve [41] and the Luikotale field site [42] from a total of four bonobo communities. We recorded all vocalisations with a 44.1 kHz sampling frequency and a 16-bit amplitude resolution with Marantz PMD 660 digital recorders and Sennheiser directional microphones (K6 power module, ME66 recording head and Rycote-Softie windscreen) at a distance of 7-10m. For each vocalisation we noted the date and ID of the caller. We defined a vocalization (or a call) as any continuous sound vocally produced by a single individual without a silent gap.

Data preparation

Only calls for which the caller was known were included in the analysis. We visually inspected spectrograms using Adobe Audition 2020 software (v. 13.0.13.46, Adobe Systems Inc., San Jose, CA, U.S.A.) with a Hamming window and a 256-frequency step. We extracted non-overlapped call units of sufficient quality (e.g., minimal background noise, no clipping) for further analysis resulting in a dataset of 1509 vocalisations from 53 adult individuals (mean: 28 calls/individual, range: 1–145 calls/individual, see S1 Table in S1 File) with broadly similar numbers of calls from female and male individuals (804 and 705, respectively). At Luikotale, individuals from one community contributed 672 calls and at Kokolopori, individuals from the Kokoalongo, Ekalakala and Fekako community contributed 526, 241 and 70 calls, respectively.

We assessed the extent to which each call could be categorised into one of 15 “original call types” derived from previously published repertoires [26,28,29]. For an overview of the sample size for each of the 15 original call types per individual, see S1 Table in S1 File.

Data analysis

Acoustic measurements. Using the R (v. 4.3.1 [43]) package warbleR [44] we used a 200 Hz high-pass filter and a 4000 Hz low-pass filter to remove low and high frequency noise (e.g., cicadas). For each call, we automatically extracted 26 time- and frequency-related acoustic parameters commonly used in bioacoustic analyses (see S2 Table in S1 File

for an exact description of all parameters and S6 Table in S1 File for a comparison of parameters used in this and other recent studies) using the package warbleR (see supplementary R code for more information). Following [37], we also calculated a pairwise distance matrix using dynamic time warping. We used classical multi-dimensional scaling (MDS) to translate the matrix into a five-dimensional space and used the axis coordinates for each sample as additional call metrics (i.e., five dynamic time wrapping MDS coordinates per call). This resulted in a total of 31 automatically extracted parameters for each call. To mitigate errors from automatic extraction of acoustic parameters whilst simultaneously retaining the calls in the dataset, we identified outliers by calculating a z-score for each acoustic parameter. Calls for which the absolute z-score was above 3.29 for a specific acoustic parameter were considered outliers: these values were replaced by the median of that acoustic parameter, following [45]. In total we identified 305 parameter values considered to be outliers from 148 calls.

Call type classification. As previously described, quantitative approaches to capture vocal repertoires range from traditional acoustic analysis and discrimination-based statistical analyses [34–36], to more state-of-the-art machine learning-based methods where supervised or unsupervised algorithms cluster calls into defined or undefined categories, respectively [29,37,38]. An unsupervised classification method (i.e., which detects underlying structure within unlabelled data) would be an objective approach and ensure a robust repertoire, but relies on substantially larger datasets. In a pilot step, we conducted an unsupervised random forest analysis, which could only categorise calls into two call types (see S3 Analysis in S1 File). Previous work on bonobo vocal behaviour (e.g., [17,20–23,26,28–31]), indicates bonobo vocalizations fall into more than two categories. As such, this result probably does not reflect acoustic differences between the calls, but is rather a consequence of the limitations of the methodology – namely a relatively small dataset for the amount of call types and the inherent gradedness of the bonobo vocal system. Since a large enough dataset necessary to implement unsupervised analyses is not available, supervised approaches are, to our knowledge, the only relevant alternative.

We followed the methods laid out in [37] for using supervised random forest analyses to assess the robustness of call classification. The random forest analysis is a machine learning method which creates a set of decision trees [46] wherein, at each node of each tree, the data is divided into two classes using a random subset of the acoustic parameters [47–49]. Each datapoint is then assigned a call type based on the category chosen by the majority of trees. Finally, the data is classified as well as possible into the given categories (the “original call types”). We used the randomForest R package [50] to implement the supervised random forest with 1000 decision trees and five (i.e., the square root of the total number of features) randomly selected acoustic parameters at each split (for more details, see [47] and [50]). Random forest classifiers are considered one of the best available classification methods, as they work better than other machine learning algorithms on small datasets and can better detect small differences between classes (see [51,52]). For this reason, this approach has been commonly and successfully used in similar studies to describe vocal repertoires (e.g., [37]). Due to the imbalanced nature of our dataset, we used a weighted random forest to compensate for sample size disparities between call types. Using the “weight” argument of the randomForest R package, each call was assigned a weight proportional to the inverse of its class’s relative frequency, ensuring that rarer call types were adequately oversampled. To estimate the significance of the overall classification we used a two-tailed binomial test with a level of chance corresponding to the number of call types to be classified (i.e., $1/15 = 0.067$).

We deemed a call type to be reliably classified, and hence acoustically discriminable, if the random forest was able to correctly classify a plurality of the calls within a given call type as the “original call type” (i.e., the count of calls labelled as the initial call type by the random forest was the highest). If the random forest incorrectly classified the plurality of calls within a given call type, we did not consider it acoustically distinct and the initial putative call type was relabelled as the call type that the random forest most often classified it as. To help clarify our approach, consider a hypothetical dataset with three initial putative call types – A, B, and C – each of which has 100 datapoints. After training a random forest model on the three putative call types, the model correctly classified 20/100 calls labelled call type A, and incorrectly classified

the other 80/100 (50/100 were classified as call type B and 30/100 were classified as call type C). The model correctly classified 100% of calls labelled call type B and call type C. According to our criterion, call types B and C would be considered valid call types, but call type A would not be considered acoustically discriminable because the model classified it as a different call type in a plurality of cases. After this result, all calls labelled call type A would be relabelled call type B because it was classified as call type B in the plurality of cases.

Data visualization. To provide an accompanying visualisation of the clustering of call types we employed a non-linear dimensionality reduction algorithm (i.e., t-Distributed Stochastic Neighbor Embedding (t-SNE)). Specifically, a t-SNE provides a visual representation of how acoustically similar calls cluster together in a 2-D space [53,54].

Results

The random forest agreed with the "original call types" on the classification of 842 of the 1509 calls (55.8% of the calls), a rate significantly exceeding the probability of a call being assigned to one of the 15 call types by chance (1/15*100 = 6.67%, binomial test p=<0.001). Accordingly, the probability of being incorrectly classified (the out-of-bag error rate) was 44.2%. The classification error varied with the call types (Table 1), with some call types being more accurately classified (e.g., *high hoots*, *screams* and *laughter*) and others less so (e.g., *barks*, *scream bark*). Eleven of the 15 call types met our criterion to be considered a reliably acoustically discriminable vocalisation (Table 2; for more information on the influence each acoustic variable had on the random forest see S4 Fig in S1 File). Follow-up t-SNE-based visualisation of the data indicates that whilst these 11 acoustically discriminable call types could be reliably identified, the bonobo vocal repertoire is still highly graded (Fig 1). We provide an overview of the 11 defined call types with an accompanying spectrogram and description of how to identify them including their the mean and range for common parameters (Fig 2).

Table 1. Confusion matrix illustrating the classification of call types by the random forest clustering analysis.

Call type	Predicted group membership															Classification error
	High hoot	*Scream*	*Grunt*	*Peep*	*Laughter*	*Low hoot*	*Whistle*	*Contest hoot*	*Pant grunt*	*Yelp*	*Peep yelp*	*Scream bark*	*Wieew bark*	*Bark*	*Soft bark*	
High hoot	**243**	1	0	0	0	0	2	3	0	4	11	1	8	9	2	0.147
Scream	14	**58**	0	3	0	0	6	0	0	0	0	1	1	2	1	0.326
Grunt	0	0	**97**	5	4	8	0	0	8	2	9	0	0	0	0	0.267
Peep	9	0	11	**102**	0	0	2	2	2	8	14	0	0	0	1	0.325
Laughter	1	0	6	0	**22**	0	0	0	0	0	0	0	0	0	0	0.241
Low hoot	2	0	22	0	0	**48**	0	0	0	1	0	0	2	0	1	0.368
Whistle	20	6	0	4	0	0	**50**	3	0	2	2	0	0	2	2	0.451
Contest hoot	6	0	0	3	0	0	1	**37**	0	1	6	2	0	1	1	0.362
Pant grunt	0	0	6	0	0	1	0	0	**36**	0	10	0	0	0	0	0.321
Yelp	21	0	12	12	0	2	2	1	1	**39**	17	0	0	0	0	0.636
Peep yelp	21	0	16	30	0	1	3	2	4	16	**58**	0	0	1	5	0.631
Scream bark	16	5	0	0	0	0	3	4	0	0	0	**9**	0	2	1	0.775
Wieew bark	34	0	2	0	1	2	0	2	0	0	2	0	**15**	0	0	0.741
Bark	49	1	1	1	0	1	6	1	2	1	15	1	0	**22**	3	0.788
Soft bark	23	0	5	4	0	0	3	7	3	3	15	0	2	8	**6**	0.924

The random forest analysis classified call types based on their acoustic features. Bold numbers in the diagonal show the absolute number of calls being classified correctly as an original call type. The non-bold cells show the absolute number of calls that were categorized as another call type. We deemed a call type to be reliably discriminated when it was consistently classified by the random forest (i.e., the count of calls labelled as the original call type by the random forest was the highest). The sample size of the "original call types" is the sum of the numbers in one row. Calls below the black line are those that are classified more as other call types than themselves and hence merged with the call type they were confused with most (see S. Fig 2 for spectrograms of these calls). The overall classification error rate was 44.2%.

https://doi.org/10.1371/journal.pone.0330250.t001

Table 2. Originally used call types and call types they were subsequently merged into, based on the random forest analysis.

Original Call type	Updated Vocal Repertoire
High hoot	*High hoot*
Scream bark	
Wieew bark	
Bark	
Soft bark	
Scream	*Scream*
Grunt	*Grunt*
Peep	*Peep*
Laughter	*Laughter*
Low hoot	*Low hoot*
Whistle	*Whistle*
Contest hoot	*Contest hoot*
Pant grunt	*Pant grunt*
Yelp	*Yelp*
Peep yelp	*Peep yelp*

https://doi.org/10.1371/journal.pone.0330250.t002

Discussion

Using supervised random forest analysis, we provide the first quantitative validation of the vocal repertoire of wild bonobos. We found that calls can be reliably discriminated into 11 previously described vocalisation types including *contest hoots, grunts, high hoots, laughters, low hoots, peeps, pant grunts, peep yelps, screams, whistles* and *yelps.*

Whilst our approach could acoustically validate many of the call types previously documented via more qualitative approaches, there were also some subtle differences. One key discontinuity with existing work, is the grouping of various *barks* (*scream barks*, *wieew barks*, *barks* and *soft barks*) and *high hoots* (with varying degrees of confidence) into a single category. Specifically, we found that all these *barks* were more often classified as *high hoots* than their own call type category, suggesting that these four previously discriminated call types rather represent one single call class, here termed *high hoots*. In their qualitative description of the vocal repertoire of bonobos, Bermejo & Omedes [28] also did not differentiate between *scream barks*, *wieew barks*, *high hoots* and *barks*, although they did distinguish between *soft barks* and *barks*.

Where findings also diverge subtly with previous work is with regards to the overall number of comprising call types: previous vocal repertoires [26,28] identified 15 and 18 call types, respectively, compared to the 11 clusters we detected. A number of factors might explain these differences. Firstly, our analysis focused exclusively on adult bonobos and hence we excluded additional calls that we recorded from other age classes, such as *pout moans*, which are primarily emitted by immatures. Furthermore, certain rare call types, such as *croaks*, *hiccups* and *moans* (described by [28]) could not be included in our dataset due to their highly infrequent occurrence which precluded quantitative analyses, again reducing the number of potential call types that could be acoustically discriminated. In a similar vein, our approach also did not allow us to unveil previously undescribed call types, and rather only verified or advocated for a merging of existing call types. Further work should follow-up on our study, using unsupervised classification algorithms on a larger sample of bonobo calls, for an even more objective classification of the bonobo vocal repertoire and to potentially describe novel call types.

Lastly, whilst the random forest analyses converged on 11 clusters, a considerable degree of acoustic overlap still exists between the call types, confirming previous research suggesting the bonobo vocal system is not discrete as is the case in other primate species such as blue monkeys [56], but rather graded, as has also been shown in Barbary

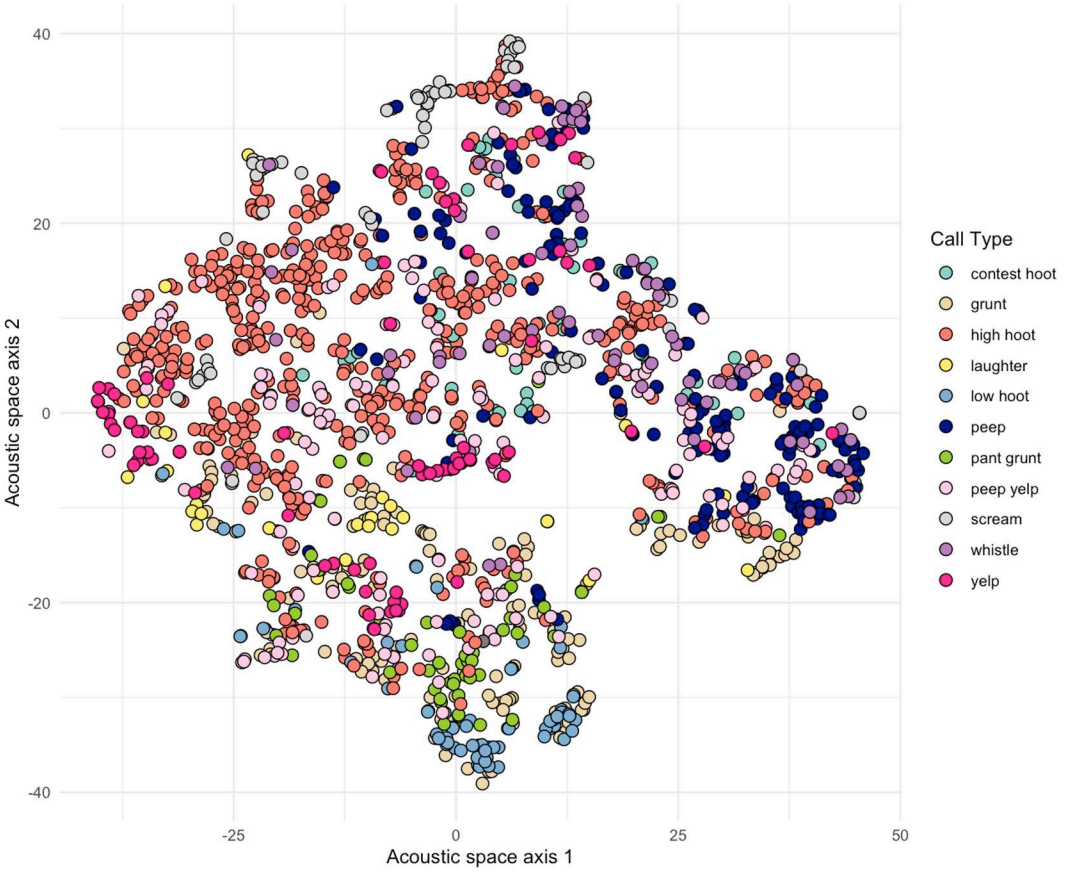
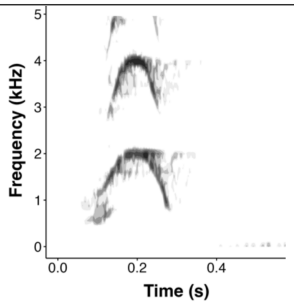
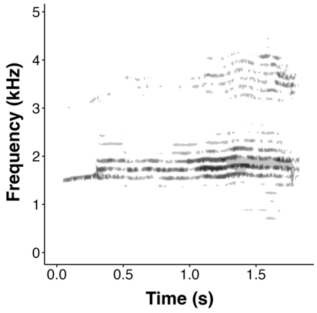
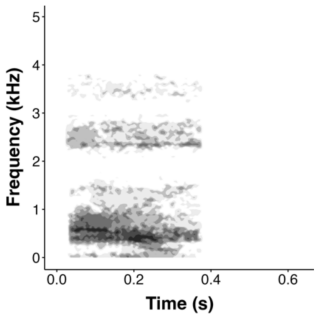
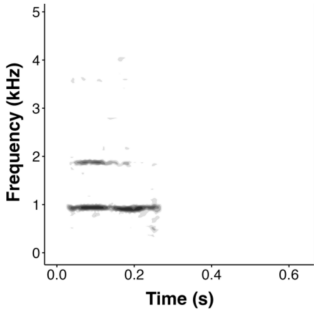


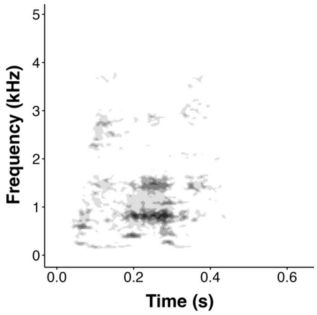
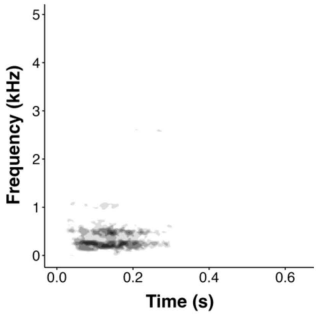
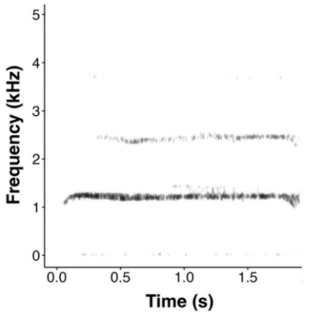
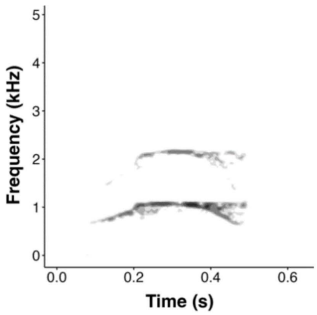
Fig 1. T-SNE scatterplot illustrating how the 11 call types from the updated vocal repertoire cluster together based on similarity in their acoustic structure. Each point in the scatterplot represents a call and the different colours of the points depict the different call types.

https://doi.org/10.1371/journal.pone.0330250.g001

macaques (*Macaca sylvanus*) [57], squirrel monkeys (*Saimiri sciureus*) [58], lemurs (*Varecia variegata*) [59], chacma baboons (*Papio ursinus*) [60] and also in non-primate vocal systems such as dolphins (*Tursiops*) [61]. Although out of the scope of the current study, follow-up research could measure and quantify the extent of gradation for each call type and for the system as a whole within a broader comparative framework. Indeed, recent work has suggested fuzzy clustering, where the degree of gradation of different call types within a repertoire can be assessed [60,62], to be one key approach that could help to further capture the precise complexity, and with it, the potential flexibility [29] of the bonobo vocal system.

Although quantitative approaches to resolving animal vocal repertoires, such as those implemented here, better avoid the subjective classification of call types, they do not come without their own shortcomings. Specifically, we encountered several obstacles that influenced the quality and analysis of the gathered data that are essential for such "data-hungry" approaches. In line with previous work on similar questions, our dataset can be characterized as SUNG (Small, Unbalanced, Noisy, but Genuine: [30]) but in addition, we arguably face an even more challenging and SUNG dataset since we compiled data from individuals in their natural habitat, the rainforest. There, visibility, inherently lower than in captivity, constrains call collection, and unavoidable background noise, including cicadas, birds, and overlapping vocalisations of other bonobos persists. Background noise in particular represented an important constraint since it made resolving the acoustic parameters for analysis inherently more challenging, ultimately reducing the number of calls available for follow-up

Call type	Spectrogram	Description
high hoot		**Peak frequency (kHz):** 2.3 ± 4.06 **Dominant frequency (kHz):** 2.03 ± 2.59 **Mean frequency (kHz):** 2.25 ± 2.51 **Shape:** inverted "U", with or without a noisy short element in the beginning **Freq. range:** high **Duration:** 0.25s ± 0.63 **Tonality:** tonal (can have scream elements) **Noise:** low
scream		**Peak frequency (kHz):** 3.54 ± 3.37 **Dominant frequency (kHz):** 3.37 ± 2.65 **Mean frequency (kHz):** 2.85 ± 1.98 **Shape:** flat and long without an inverted "U" at start or end **Freq. range:** high **Duration:** 0.58s ± 1.24 **Tonality:** broadband **Noise:** variable
grunt		**Peak frequency (kHz):** 0.75 ± 2.85 **Dominant frequency (kHz):** 0.9 ± 2.64 **Mean frequency (kHz):** 1.59 ± 1.78 **Shape:** unshaped **Freq. range:** small, very low frequency **Duration:** 0.1s ± 0.28 **Tonality:** tonal or atonal **Noise:** high **Other:** can sound "breathy"
peep		**Peak frequency (kHz):** 1.85 ± 3.68 **Dominant frequency (kHz):** 1.97 ± 3.19 **Mean frequency (kHz):** 2.03 ± 1.77 **Shape:** flat, no or one harmonic, sometimes only with a small squiggle or upwards slope in the frequency **Freq. range:** medium to high **Duration:** 0.14 s± 0.47 **Tonality:** tonal **Noise:** low

laughter		**Peak frequency (kHz)**: **0.87 ± 4.06** **Dominant frequency (kHz)**: **1.12 ± 2.45** **Mean frequency (kHz)**: **1.76 ± 0.88** **Shape:** unshaped **Freq. range:** medium **Duration: 0.14s ± 0.36** **Tonality:** atonal **Noise:** high
low hoot		**Peak frequency (kHz):** 0.55 ± 1.12 **Dominant frequency (kHz):** 0.63 ± 0.89 **Mean frequency (kHz):** 1.2 ± 1.31 **Shape:** unshaped **Freq. range:** small **Duration:** 0.12s ± 0.43 **Tonality:** variable **Noise:** high
whistle		**Peak frequency (kHz):** 2.18 ± 4.06 **Dominant frequency (kHz):** 2.27 ± 1.9 **Mean frequency (kHz):** 2.36 ± 1.86 **Shape:** flat or with “squiggles” **Freq. range:** low but high pitched **Duration:** 0.5 ± 1.47 **Tonality:** tonal **Noise:** low
contest hoot		**Peak frequency (kHz):** 2.32 ± 1.98 **Dominant frequency (kHz):** 2.24 ± 1.46 **Mean frequency (kHz):** 2.31 ± 1.2 **Shape:** inverted “U” with same frequency at beginning and end **Freq. range:** low **Duration:** 0.24s ± 0.32 **Tonality:** tonal **Noise:** low

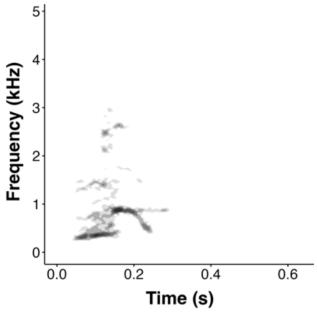
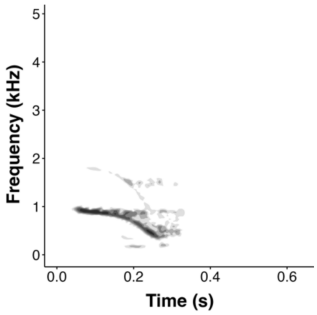
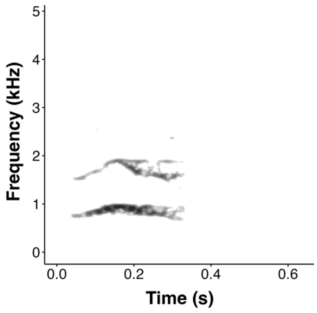
pant grunt		**Peak frequency (kHz)**: **0.86 ± 1.98** **Dominant frequency (kHz)**: **0.99 ± 0.88** **Mean frequency (kHz)**: **1.44 ± 0.76** **Shape**: upwards slope **Freq. range:** high **Duration: 0.08s ± 0.12** **Tonality:** tonal **Noise:** medium
yelp		**Peak frequency (kHz):** 1.78 ± 3.97 **Dominant frequency (kHz):** 1.85 ± 1.5 **Mean frequency (kHz):** 2.04 ± 1.83 **Shape:** downwards pointing slope **Freq. range:** low **Duration:** 0.15s ± 0.48 **Tonality:** tonal **Noise:** low
peep yelp		**Peak frequency (kHz):** 1.55 ± 4.06 **Dominant frequency (kHz):** 1.55 ± 2.2 **Mean frequency (kHz):** 1.96 ± 1.72 **Shape:** variable, inverted "u" or upwards pointing **Freq. range:** low **Duration:** 0.14s ± 0.23 **Tonality:** tonal **Noise:** low

Fig 2. Spectrogram and description of the eleven call types of the updated bonobo vocal repertoire. Note that the x-axes of the spectrograms vary in length. For each call type, mean and range for duration; peak, dominant and mean frequency are given. Additionally, qualitative identification characteristics are used to describe each call type: "Shape" refers to the form of the fundamental frequency and/or harmonics; "Freq. range" refers to the frequency range of the calls; "Tonality" to the harmonicity of the call; "Noise" to the signal to noise ratio and, if applicable, "Other" refers to additional noteworthy identification characteristics. Spectrograms were made in R with the dynaSpec package [55] with a hanning window, the minimum decibel to be included in the spectrogram set at −30 and silent margins were added at beginning and end. For the spectrograms, corresponding audio recordings can be found in the Electronic Supplementary Material (S2–S12). Spectrograms for the four "bark" variants that are now all classed as "high hoots" can be found in S5 Fig in S1 File.

https://doi.org/10.1371/journal.pone.0330250.g002

acoustic analyses. A further obstacle encountered in this study was the mandatory 7m distance between humans and bonobos that needed to be adhered to, to reduce disease transmission and avoid overly-interfering with bonobo behaviour [39]. This has little influence on higher amplitude long-distance calls such as *high hoots*, *contest hoots* and *whistles*; however, it impacted our capacity to record good quality soft, short-distance calls, including *pant grunts* and *grunts*, again

representing an additional bottleneck on recordings available for quantification. In addition, our definition for reliably identifying a call type whilst somewhat arbitrary, was motivated by the underlying assumption that for a call type to be reliably discriminated, the random forest needed to correctly classify the plurality of calls. We acknowledge the limitations of this study and hope future studies will validate the here-presented repertoire, particularly when using varying quantitative approaches and threshold values.

Despite these issues, we are confident our study represents an important milestone at quantitatively resolving the vocal repertoire of wild bonobos. We hope this work will catalyse similar studies in other nonhuman species where more objective repertoires are still missing, including, and particularly surprisingly, chimpanzees. Future work leveraging an even larger sample size could also consider extending our approach to include more unsupervised machine learning-based approaches. Such methods are arguably even more objective since calls are categorized independently of pre-existing call categories, further removing observer bias inherent to more supervised approaches.

Lastly, this study provides essential groundwork for follow-up quantitative investigations into the contexts accompanying call types. In particular, a whole repertoire approach can now be adopted to probe how bonobo call types are associated with specific social and environmental events and what light this can shed on their underlying function (see for example [63]). Such follow-up work will ultimately allow for a more detailed and holistic understanding of bonobo communication.

Supporting information

S1 File. S1 Table. Sample size of the original call types per individual. S2 Table. Acoustic parameters used in the random forest analysis. The abbreviation of the parameter (which is used in S1 Fig), the full name and a description of the parameter are given. This table is partly taken from (Keen et al. 2021). S3 Analysis. Unsupervised random forest approach. S4 Fig. Influence (mean decrease accuracy) each acoustic parameter has on the random forest model. S5 Fig. Spectrograms of four variants of the "high hoot" call type. These calls were formerly, as "original call types", categorized as "soft bark", "bark", "wieew bark" and "scream bark". Using our random forest analysis, these former call types are now merged together into a single call category: "high hoots". S6 Table. Acoustic parameter comparison. The two hitherto performed quantitative analyses of a subset of calls of the vocal repertoire of bonobos ([30] and [29]) and our study used broadly similar, commonly used acoustic parameters for the quantitative acoustic analyses. All three studies used parameters such as duration and parameters related to the distribution of the energy within the call. Whilst we in the study at hand analysed acoustic parameters related to the dominant frequency, Arnaud et al. [30] and Keenan et al. [29] used acoustic features regarding the fundamental frequency. Oftentimes, but not always, the dominant frequency correlates highly with the fundamental frequency. In addition, whilst we used dynamic time warping-related parameters, Arnaud et al. used MFCCs, and Keenan used neither of the two.
(ZIP)

S2 File. high hoot. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S3 File. scream. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S4 File. grunt. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S5 File. peep. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S6 File. laughter. Corresponding audio recording for the spectrogram Fig 2.
(WAV)

S7 File. low hoot. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S8 File. whistle. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S9 File. contest hoot. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S10 File. pant grunt. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S11 File. yelp. Corresponding audio recording for the spectrogram in Fig 2.
(WAV)

S12 File. peep yelp.Corresponding audio recording for the spectrogram in Fig 2. .
(WAV)

S1 Data. S7 Questionnaire.
(DOCX)

Acknowledgments

We thank the pisteurs of the Kokolopori Bonobo Project for their invaluable help with data collection, the Institut Congolais pour la Conservations de la Nature (ICCN) and the Ministry of Scientific Research and Technology in the DRC for their permission to work in the Democratic Republic of the Congo and the Bonobo Conservation Initiative and Vie Sauvage for support. We thank the LuiKotale Bonobo Project for offering access to the study site, the ICCN for granting permission to conduct research in the buffer zone of Salonga National Park, and the people of Lompole village for hosting researchers in their forest. MB thanks Lara Zanutto for help with data collection. IS thanks Claudia Wilke for help with extracting calls. We thank Tim Sainburg for helpful methodological discussions and Nikola Falk for sharing her code to loop the dynaSpec package to create the spectrograms.

Author contributions

Conceptualization: Franziska Wegdell, Simon W. Townsend.

Data curation: Franziska Wegdell, Isaac Schamberg, Mélissa Berthet.

Formal analysis: Franziska Wegdell, Isaac Schamberg, Mélissa Berthet.

Funding acquisition: Martin Surbeck, Simon W. Townsend.

Investigation: Franziska Wegdell, Isaac Schamberg, Mélissa Berthet, Simon W. Townsend.

Methodology: Franziska Wegdell, Isaac Schamberg, Mélissa Berthet, Simon W. Townsend.

Project administration: Franziska Wegdell, Simon W. Townsend.

Resources: Martin Surbeck, Simon W. Townsend.

Software: Franziska Wegdell, Isaac Schamberg, Mélissa Berthet.

Supervision: Martin Surbeck, Simon W. Townsend.

Validation: Franziska Wegdell, Isaac Schamberg, Mélissa Berthet, Yannik Rothacher, Volker Dellwo, Simon W. Townsend.

Visualization: Franziska Wegdell, Isaac Schamberg, Mélissa Berthet.

Writing – original draft: Franziska Wegdell, Simon W. Townsend.

Writing – review & editing: Franziska Wegdell, Isaac Schamberg, Mélissa Berthet, Martin Surbeck, Simon W. Townsend.

References

1. Crockford C. Why does the chimpanzee vocal repertoire remain poorly understood? - and what can be done about it. The chimpanzees of the Taï forest: 40 years of research. Cambridge University Press; 2019. pp. 394–409.
2. Mitani J, Gros-Louis J. Chorusing and call convergence in Chimpanzees: tests of three hypotheses. Behav. 1998;135(8):1041–64. https://doi.org/10.1163/156853998792913483
3. Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, et al. Vocal learning in the functionally referential food grunts of chimpanzees. Curr Biol. 2015;25(4):495–9. https://doi.org/10.1016/j.cub.2014.12.032 PMID: 25660548
4. Crockford C, Wittig RM, Mundry R, Zuberbühler K. Wild chimpanzees inform ignorant group members of danger. Curr Biol. 2012;22(2):142–6. https://doi.org/10.1016/j.cub.2011.11.053 PMID: 22209531
5. Girard-Buttoz C, Surbeck M, Samuni L, Tkaczynski P, Boesch C, Fruth B, et al. Information transfer efficiency differs in wild chimpanzees and bonobos, but not social cognition. Proc Biol Sci. 2020;287(1929):20200523. https://doi.org/10.1098/rspb.2020.0523 PMID: 32576115
6. Kalan AK, Boesch C. Audience effects in chimpanzee food calls and their potential for recruiting others. Behav Ecol Sociobiol. 2015;69(10):1701–12. https://doi.org/10.1007/s00265-015-1982-1
7. Schel AM, Townsend SW, Machanda Z, Zuberbühler K, Slocombe KE. Chimpanzee alarm call production meets key criteria for intentionality. PLoS One. 2013;8(10):e76674. https://doi.org/10.1371/journal.pone.0076674 PMID: 24146908
8. Townsend SW, Koski SE, Byrne RW, Slocombe KE, Bickel B, Boeckle M, et al. Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals. Biol Rev Camb Philos Soc. 2017;92(3):1427–33. https://doi.org/10.1111/brv.12289 PMID: 27480784
9. Townsend SW, Deschner T, Zuberbühler K. Female chimpanzees use copulation calls flexibly to prevent social competition. PLoS One. 2008;3(6):e2431. https://doi.org/10.1371/journal.pone.0002431 PMID: 22423311
10. Bortolato T, Friederici AD, Girard-Buttoz C, Wittig RM, Crockford C. Chimpanzees show the capacity to communicate about concomitant daily life events. iScience. 2023;26(11):108090. https://doi.org/10.1016/j.isci.2023.108090 PMID: 37876805
11. Crockford C, Boesch C. Call combinations in wild chimpanzees. Behaviour. 2005.
12. Girard-Buttoz C, Zaccarella E, Bortolato T, Friederici AD, Wittig RM, Crockford C. Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. Commun Biol. 2022;5(1):410. https://doi.org/10.1038/s42003-022-03350-8 PMID: 35577891
13. Girard-Buttoz C, Neumann C, Bortolato T, Zaccarella E, Friederici AD, Wittig RM. Versatile use of chimpanzee call combinations promotes meaning expansion. Sci Adv. 2025;11(19):eadq2879. https://doi.org/10.1126/sciadv.adq2879 PMID: 40344055
14. Leroux M, Bosshard AB, Chandia B, Manser A, Zuberbühler K, Townsend SW. Chimpanzees combine pant hoots with food calls into larger structures. Anim Behav. 2021;179:41–50. https://doi.org/10.1016/j.anbehav.2021.06.026
15. Leroux M, Schel AM, Wilke C, Chandia B, Zuberbühler K, Slocombe KE, et al. Call combinations and compositional processing in wild chimpanzees. Nat Commun. 2023;14(1):2225. https://doi.org/10.1038/s41467-023-37816-y PMID: 37142584
16. Girard-Buttoz C, Bortolato T, Laporte M, Grampp M, Zuberbühler K, Wittig RM, et al. Population-specific call order in chimpanzee greeting vocal sequences. iScience. 2022;25(9):104851. https://doi.org/10.1016/j.isci.2022.104851 PMID: 36034222
17. Genty E, Clay Z, Hobaiter C, Zuberbühler K. Multi-modal use of a socially directed call in bonobos. PLoS One. 2014;9(1):e84738. https://doi.org/10.1371/journal.pone.0084738 PMID: 24454745
18. Berthet M, Surbeck M, Townsend SW. Extensive compositionality in the vocal system of bonobos. Science. 2025;388(6742):104–8. https://doi.org/10.1126/science.adv1170 PMID: 40179186
19. Schamberg I, Cheney DL, Clay Z, Hohmann G, Seyfarth RM. Call combinations, vocal exchanges and interparty movement in wild bonobos. Anim Behav. 2016;122:109–16. https://doi.org/10.1016/j.anbehav.2016.10.003
20. Schamberg I, Cheney DL, Clay Z, Hohmann G, Seyfarth RM. Bonobos use call combinations to facilitate inter-party travel recruitment. Behav Ecol Sociobiol. 2017;71(4). https://doi.org/10.1007/s00265-017-2301-9
21. Clay Z, Zuberbühler K. Food-associated calling sequences in bonobos. Anim Behav. 2009;77(6):1387–96. https://doi.org/10.1016/j.anbehav.2009.02.016
22. Clay Z, Zuberbühler K. Bonobos extract meaning from call sequences. PLoS One. 2011;6(4):e18786. https://doi.org/10.1371/journal.pone.0018786 PMID: 21556149
23. Clay Z, Archbold J, Zuberbühler K. Functional flexibility in wild bonobo vocal behaviour. PeerJ. 2015;3:e1124. https://doi.org/10.7717/peerj.1124 PMID: 26290789

24. Schamberg I, Clay Z, Townsend SW, Surbeck M. Between-group variation in production of pant-grunt vocalizations by wild bonobos (*Pan paniscus*). Behav Ecol Sociobiol. 2023;77(1). https://doi.org/10.1007/s00265-022-03285-4
25. Schamberg I, Surbeck M, Townsend SW. Cross-population variation in usage of a call combination: evidence of signal usage flexibility in wild bonobos. Anim Cogn. 2024;27(1):58. https://doi.org/10.1007/s10071-024-01884-4 PMID: 39212694
26. De Waal FBM. The communicative repertoire of captive Bonobos (*Pan paniscus*), Compared To That of Chimpanzees. Behaviour. 1988;106(3–4):183–251. https://doi.org/10.1163/156853988x00269
27. Hopkins WD, Savage-Rumbaugh ES. Vocal communication as a function of differential rearing experiences in *Pan paniscus*: a preliminary report. Int J Primatol. 1991;12(6):559–83. https://doi.org/10.1007/bf02547670
28. Bermejo M, Omedes A. Preliminary vocal repertoire and vocal communication of wild bonobos (*Pan paniscus*) at Lilungu (Democratic Republic of Congo). Folia Primatol (Basel). 1999;70(6):328–57. https://doi.org/10.1159/000021717 PMID: 10640882
29. Keenan S, Mathevon N, Stevens JMG, Nicolè F, Zuberbühler K, Guéry J-P, et al. The reliability of individual vocal signature varies across the bonobo's graded repertoire. Anim Behav. 2020;169:9–21. https://doi.org/10.1016/j.anbehav.2020.08.024
30. Arnaud V, Pellegrino F, Keenan S, St-Gelais X, Mathevon N, Levréro F, et al. Improving the workflow to crack Small, Unbalanced, Noisy, but Genuine (SUNG) datasets in bioacoustics: the case of bonobo calls. PLoS Comput Biol. 2023;19(4):e1010325. https://doi.org/10.1371/journal.pcbi.1010325 PMID: 37053268
31. Hohmann G, Fruth B. Structure and use of distance calls in wild bonobos (*Pan paniscus*). Int J Primatol. 1994;15(5):767–82. https://doi.org/10.1007/bf02737430
32. Sills JM, Reichmuth C. Vocal behavior in spotted seals (*Phoca largha*) and implications for passive acoustic monitoring. Front Remote Sens. 2022;3. https://doi.org/10.3389/frsen.2022.862435
33. Crunchant A, Borchers D, Kühl H, Piel A. Listening and watching: Do camera traps or acoustic sensors more efficiently detect wild chimpanzees in an open habitat? Methods Ecol Evol. 2020;11(4):542–52. https://doi.org/10.1111/2041-210x.13362
34. Hending D, Seiler M, Stanger-Hall KF. The vocal repertoire of the Northern Giant Mouse Lemur (*Mirza zaza*) in Captivity. Int J Primatol. 2020;41(5):732–63. https://doi.org/10.1007/s10764-020-00181-4
35. Maretti G, Sorrentino V, Finomana A, Gamba M, Giacoma C. Not just a pretty song: an overview of the vocal repertoire of *Indri indri*. J Anthropol Sci. 2010;88:151–65. PMID: 20834055
36. Salmi R, Hammerschmidt K, Doran-Sheehy DM. Western Gorilla Vocal repertoire and contextual use of vocalizations. Ethology. 2013;119(10):831–47. https://doi.org/10.1111/eth.12122
37. Keen SC, Odom KJ, Webster MS, Kohn GM, Wright TF, Araya-Salas M. A machine learning approach for classifying and quantifying acoustic diversity. Methods Ecol Evol. 2021;12(7):1213–25. https://doi.org/10.1111/2041-210x.13599 PMID: 34888025
38. Sainburg T, Thielk M, Gentner TQ. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. PLoS Comput Biol. 2020;16(10):e1008228. https://doi.org/10.1371/journal.pcbi.1008228 PMID: 33057332
39. Gilardi KV, Gillespie TR, Leendertz FH, Macfie EJ, Travis DA, Whittier CA. Best Practice Guidelines for Health Monitoring and Disease Control in Great Ape Populations. Gland, Switzerland: IUCN SSC Primate Specialist Group; 2015. pp. 56. www.iucn.org/what/work_by_topic
40. Altmann J. Observational study of behavior: sampling methods. Behaviour. 1974;49(3):227–67. https://doi.org/10.1163/156853974x00534 PMID: 4597405
41. Surbeck M, Coxe S, Lokasola AL. Lonoa: The Establishment of a Permanent Field Site for Behavioural Research on Bonobos in the Kokolopori Bonobo Reserve. Pan Afr News. 2017;24(2):13–5. https://doi.org/10.5134/228898
42. Hohmann G, Fruth B. Lui Kotal - A new site for field research on bonobos in the Salonga National Park. Pan Afr News. 2003;10(2):25–7. https://doi.org/10.5134/143430
43. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2023. Available from: https://www.r-project.org/
44. Araya-Salas M, Smith-Vidaurre G. warbleR: an r package to streamline analysis of animal acoustic signals. Methods Ecol Evol. 2016;8(2):184–91. https://doi.org/10.1111/2041-210x.12624
45. Santhanam T, Padmavathi MS. Comparison of K-Means clustering and statistical outliers in reducing medical datasets. In: 2014 International Conference on Science Engineering and Management Research (ICSEMR), 2014. pp. 1–6. https://doi.org/10.1109/icsemr.2014.7043602
46. Kotsiantis SB. Decision trees: a recent overview. Artif Intell Rev. 2011;39(4):261–83. https://doi.org/10.1007/s10462-011-9272-4
47. Breimann L. Random Forests. Mach Learn. 2001;12343 LNCS:503–515. https://doi.org/10.1007/978-3-030-62008-0_35
48. Rothacher Y, Strobl C. Identifying Informative Predictor Variables With Random Forests. J Educ Behav Stat. 2023;49(4):595–629. https://doi.org/10.3102/10769986231193327
49. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol Methods. 2009;14(4):323–48. https://doi.org/10.1037/a0016973 PMID: 19968396
50. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2:18–22.
51. Carugati F, Friard O, Protopapa E, Mancassola C, Rabajoli E, De Gregorio C, et al. Discrimination between the facial gestures of vocalising and non-vocalising lemurs and small apes using deep learning. Ecol Inform. 2025;85:102847. https://doi.org/10.1016/j.ecoinf.2024.102847

52. Wierucka K, Murphy D, Watson SK, Falk N, Fichtel C, León J, et al. Same data, different results? Machine learning approaches in bioacoustics. Methods Ecol Evol. 2025;16(8):1574–86. https://doi.org/10.1111/2041-210x.70091
53. van der Maaten LJP. Accelerating t-SNE using Tree-Based Algorithms. J Mach Learn Res. 2014;15:3221–45.
54. van der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-SNE. J Mach Learn Res. 2008;9: 2579–605.
55. Araya-Salas M, Wilkins MR. dynaSpec: dynamic spectrogram visualizations in R. 2020.
56. Fuller JL. The vocal repertoire of adult male blue monkeys (*Cercopithecus mitis* stulmanni): a quantitative analysis of acoustic structure. Am J Primatol. 2014;76(3):203–16. https://doi.org/10.1002/ajp.22223 PMID: 24130044
57. Hammerschmidt K, Fischer J. The Vocal Repertoire of *Barbary Macaques*: A Quantitative Analysis of a Graded Signal System. Ethology. 1998;104(3):203–16. https://doi.org/10.1111/j.1439-0310.1998.tb00063.x
58. Winter P, Ploog D, Latta J. Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. Exp Brain Res. 1966;1(4):359–84. https://doi.org/10.1007/BF00237707
59. Batist CH, Razafindraibe MN, Randriamanantena F, Baden AL. Bioacoustic characterization of the black-and-white ruffed lemur (*Varecia variegata*) vocal repertoire. Primates. 2023;64(6):621–35. https://doi.org/10.1007/s10329-023-01083-8 PMID: 37584832
60. Wadewitz P, Hammerschmidt K, Battaglia D, Witt A, Wolf F, Fischer J. Characterizing vocal repertoires--hard vs. soft classification approaches. PLoS One. 2015;10(4):e0125785. https://doi.org/10.1371/journal.pone.0125785 PMID: 25915039
61. Jones B, Zapetis M, Samuelson MM, Ridgway S. Sounds produced by bottlenose dolphins (*Tursiops*): a review of the defining characteristics and acoustic criteria of the dolphin vocal repertoire. Bioacoustics. 2019;29(4):399–440. https://doi.org/10.1080/09524622.2019.1613265
62. Cusano DA, Noad MJ, Dunlop RA. Fuzzy clustering as a tool to differentiate between discrete and graded call types. JASA Express Lett. 2021;1(6):061201. https://doi.org/10.1121/10.0005111 PMID: 36154369
63. Berthet M, Coye C, Dezecache G, Kuhn J. Animal linguistics: a primer. Biol Rev Camb Philos Soc. 2023;98(1):81–98. https://doi.org/10.1111/brv.12897 PMID: 36189714