Manuscript

Click here to access/download;Manuscript;Revised manuscript.docx

Click here to view linked References

A machine learning approach to infant distress calls and maternal behaviour of wild chimpanzees

Guillaume Dezecache[a,b,c,d,*], Klaus Zuberbühler[a,b,e], Marina Davila-Ross[c] & Christoph D. Dahl[a,f,g,*]

[a]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland;

[b]Budongo Conservation Field Station, Masindi, Uganda;

[c]Department of Psychology, University of Portsmouth, Portsmouth, England, United Kingdom;

[d]Université Clermont Auvergne, CNRS, LAPSCO, Clermont-Ferrand, France;

[e]School of Psychology and Neuroscience, University of St Andrews, St Andrews, Scotland, United Kingdom;

[f]Graduate Institute of Mind, Brain and Consciousness, Taipei Medical University, Taipei, Taiwan;

[g]Brain and Consciousness Research Center, Taipei Medical University Shuang-Ho Hospital, New Taipei City, Taiwan

[*]Correspondence:

Guillaume Dezecache <guillaume.dezecache@gmail.com> ; +33473406458; LAPSCO-UMR -CNRS 6024, 17 Rue Paul Collomp, 63000 Clermont-Ferrand, France

Christoph D. Dahl <christoph.d.dahl@gmail.com>; <christoph.dahl@tmu.edu.tw>

Acknowledgments

We thank the Uganda Wildlife Authority (UWA) and the Uganda National Council for Science and Technology (UNCST) for permission to conduct the study, and Geoffrey Muhanguzi, Caroline Asiimwe, Sam Adue and Monday M'Botella for their support in the field. We are grateful to two anonymous Reviewers, the Editor and Bahar Tuncgenc for helpful comments on the manuscript.

Abstract

Distress calls are an acoustically variable group of vocalizations ubiquitous in mammals and other animals. Their presumed function is to recruit help, but there has been much debate on whether the nature of the disturbance can be inferred from the acoustics of distress calls. We used machine learning to analyse episodes of distress calls of wild infant chimpanzees. We extracted exemplars from those distress call episodes and examined them in relation to the external event triggering them and the distance to the mother. In further steps, we tested whether the acoustic variants were associated with particular maternal behaviours. Our results suggest that, although infant chimpanzee distress calls are highly graded, they can convey information about discrete problems experienced by the infant and about distance to the mother, which in turn may help guide maternal parenting decisions. The extent to which mothers rely on acoustic cues alone (versus integrate other contextual-visual information) to decide upon intervening should be the focus of future research.

Keywords: semantics, crying, whimpers, *Pan troglodytes*, support vector machine, machine learning

Declarations

Funding

We are grateful to the Royal Zoological Society of Scotland for providing core funding to the Budongo Conservation Field Station. This research was supported by funding from the European Union's Seventh Framework Programme for research, technological development and demonstration (grant agreement no 283871), a Fyssen Foundation post-doctoral fellowship awarded to GD, the Swiss National Science Foundation (PZ00P3_154741) and Start up-funding of the Taipei Medical University (108-6402-004-112) awarded to CDD.

Authors' contributions

GD, KZ and MDR designed the research; GD collected the data; GD analysed the behavioural data; CDD implemented the acoustic extraction and the machine learning procedure; all Authors contributed to the writing of the manuscript.

Conflict of interests

None.

Code availability

Code is available at: https://github.com/ChristophDahl/Chimpanzee-Distress-Calls

Introduction

Distress calls are the most primitive mammalian vocalizations (MacLean 1985; Newman 2007). They appear early in ontogeny (Illingworth 1955) and are highly preserved in phylogeny, with a simple structure (tonal sound with a chevron or descending shape) also present in species beyond the class Mammalia (Lingle et al. 2012). Their function is most likely to recruit help (Soltis 2004; Lingle et al. 2012; Lingle and Riede 2014), but it is uncertain whether caregivers can base their reactions and inferences about the nature of the disturbance on the acoustics of the calls alone, or whether they must also rely on contextual cues.

In humans, at least three types of distress vocalization elicitors have been delineated: after birth, human infants produce cries when hungry (Gilbert and Robb 1996), when separated from their main caregiver (Christensson et al. 1995), and when in pain (Fuller 1991). Research consisting of testing whether caregivers and experienced professionals can differentiate between cries associated with a diversity of demands and respond accordingly has produced inconclusive results (Wasz-Höckert et al. 1964; Müller et al. 1974). Overall, the current view is that, in humans and other mammals, distress calls represent an acoustically graded system, with no clear acoustic boundaries (Zeskind et al. 1985; Porter et al. 1986; Lingle et al. 2012). However, the fact that a signal is graded does not automatically disqualify it from categorical perception (May et al. 1989; Fischer 1998; Green et al. 2020).

In this study, we used machine learning to identify infant needs based on acoustic variants of their distress calling. We also examined whether the distress calls can be associated with contingent maternal behaviours. We focused our analysis on our closest living relatives, the chimpanzees.

In chimpanzees, distress calls (typically: ‘whimpers’) are mainly produced by infants (Plooij et al. 1984; Goodall 1986; Bard 2000). Distress calls are different in form and function from alarm calls (typically alarm hoos and barks; see Crockford et al. 2012, 2017, 2018; Schel et al. 2013).

This difference is already potent in infants and juveniles (Dezecache et al. 2019). One useful functional distinction (at least in chimpanzees) is that distress calls attract social partners to one's particular situation, whereas alarm calls are used to warn social partners about a common danger in the environment (Dezecache et al., 2019).

Infant chimpanzee distress calls are short tonal and low-pitched sounds, which can be given in series (see Electronic Supplementary Material [ESM] for video examples; see ESM Figure 1 for spectrogram), potentially reflecting a difference in arousal and other internal states (Briefer 2012). After previous longitudinal research with free-ranging chimpanzees (Plooij et al. 1984), rough distinctions (based on ear) have been made between (a) whimpers (sequences of relatively pure tones), (b) whimper-hoos (soft and low-pitched sounds rarely produced in sequences), and (c) 'crying' (loud vocalizations marked by rapid fluctuations in frequency, resembling adult screams). Apart from work on idiosyncratic distress calls (with features of the fundamental frequency contributing to individual distinctiveness, see Levréro and Mathevon 2013), no acoustic analyses have been conducted on these vocalisations in chimpanzees. Distress calls thus seem to be a rather broad yet distinct acoustic phenomenon in young chimpanzees.

Recent developments in machine learning have proved critical in the study of animal vocal communication (Mielke and Zuberbühler 2013; Fedurek et al. 2016; Turesson et al. 2016; Versteegh et al. 2016) and human crying (Barajas-Montiel and Reyes-Garcia 2006; Saraswathy et al. 2012; Chang and Li 2016), but classification based on machine learning algorithms has not yet been applied to distress calls of chimpanzees. We analysed distress call episodes from infants of a cohort of wild chimpanzees (*Pan troglodytes schweinfurthii*) (N=8) in the Sonso community of Budongo Forest, Uganda. We extracted and analysed acoustical information from a total of 178 distress call episodes. We subjected acoustic exemplars from a recording sequence to an automated feature extraction algorithm before training a supervised learning algorithm for subsequent categorisation. The procedure consisted of training a model to segregate exemplars given in a particular context. This model was then used to categorise new exemplars (Mohri et al. 2018), which enabled us to evaluate whether the acoustics of graded distress call series encode information about the context of emission. Following prior reports in other primates, we also evaluated whether the distress calls encoded information about the distance between the infant and the mother (Bayart et al. 1990; Wiener et al. 1990). Finally, we looked at whether contingent maternal behaviour could, in principle, be predicted from the acoustic exemplars alone (although other cues are also inevitably used by the mother).

With the dual perspective from the sender and the receiver, we looked for evidence of a common acoustic code potentially used by the chimpanzee infants and mothers that could contribute to coordinate their activities, notably when parental protective behaviour is most needed.

Methods

Ethical note

We received permission from the Uganda Wildlife Authority (UWA) and the Uganda National Council for Science and Technology (UNCST) to conduct this study.

Subjects

Data were collected from infants (N=8) of the well-habituated wild chimpanzees (N≈70) of the Sonso community (Reynolds 2005) of Budongo Forest, Uganda, during February-June 2014, December 2014, March-June 2015 and April-May 2016 (see Table 1 for details). For further information about the study site, see Eggeling (1947) and Reynolds (2005). We collected data from those specific infants because they could easily be followed in the forest, as they were born from well-habituated mothers, unwary of human observers.

Data collection

Distress calling episodes were continuously videotaped in all-occurrence sampling (Altmann 1974). Despite their acoustic variability, the short, tonal, and sometimes sequential production pattern of distress calls enables trained listeners to clearly identify these calls as distinct from other infant vocalizations such as grunts or barks (see ESM Figure 1 for a spectrogram and ESM for video and audio examples). The entire dataset, part of which was used in this study, was gathered by the first author and is composed of 271 call episodes (i.e., a series of distress calls interspaced by no more than 10s). They were video-recorded using a Panasonic HC X909/V700 video-camera, with the sound captured by a Sennheiser MKE-400 shotgun microphone. To account for any missing information in the videos, all potential causes of infant distress calling, as well as other contextual information, were recorded.

Infants' perceived needs were classified according to so-called 'situations' and more specific 'problems' (see ESM Table 1 for details) using a typology established in the field. Physical distance between mother and infant was determined at the onset of the distress call episodes. We coded 'supported' if most of the weight of the infant was then supported by the mother; 'contact' if there was physical contact with mother but without full support (e.g., standing on the ground); 'arms-reach' if the infant was within the arm's reach of the mother, and 'beyond' if the infant was beyond the arm's reach of the mother. Classification of the distances in this manner proved to be more meaningful and probably reliable than the estimation of distance using a metric system (such as measurement in meters or inches). We also determined the nature of the mothers' reaction starting from the onset of the calling episode until up to 10 seconds after the offset of the calling episode. We coded for whether the mother gazed towards the infant (based on facial orientation), approached the infant, collected the infant or vocalized. Note that these are not mutually exclusive categories in that mothers may, for instance, approach and call simultaneously. While the coding of 'situations' and 'problems' largely relied on live commentaries by the cameraman, the coding of distance and maternal behaviours was mostly based on video. The coding of distance and maternal behaviours was intra-reliably validated on 19.5% of the entire dataset (κ configuration of responses = 0.93; κ distance = 1; κ gaze = 0.90; κ approach = 0.87; κ collection = 0.96; κ vocalization = 0.92).

Quantification and statistical analysis

First, we pre-processed the raw audio files (wav format, 44100 Hz) by applying a band pass filter (filter order 10) from 100 to 700 Hz. We also filtered out background noise using a Vuvuzela denoising algorithm, implementing noise spectrum extraction, signal-to-noise ratio estimation, attenuation map and inverse short-time Fourier transform computation (Boll 1979; Ephraim and Malah 1984). The dedicated MATLAB functions are freely available at MathWorks FileExchange (Choqueuse 2020). Distress calls could be above 700 Hz but this was the maximal value of the peak fundamental frequency at mid-call of the first distress call of the episodes composing the entire dataset. Despite being very conservative, this pass band was appropriate for the localization of exemplars and to avoid false positives. We subdivided the audio files into 10ms segments, determined the prominence of each exemplar time element's energy and applied a cut-off threshold at the 80th percentile of the energy-prominence distribution. We then sorted out those extractions that were shorter than 40ms or longer than 120ms (corresponding to the range of durations extracted manually from distress calls, based on the first call of the distress call episodes of the entire dataset) and subjected the remaining extractions to a human-based validation process, leaving 1330 exemplar on- and offsets. We extracted the full frequency spectrum ranging from 50 to 4000 Hz of a given exemplar on- and offsets.

We conducted a four-stage analysis, consisting of 'feature extraction', 'feature selection', 'classification' and 'feature evaluation':

Feature extraction

Feature extraction is the extraction of a subset of relevant features from each exemplar to minimize data size, redundancy in information and computational efforts and increase generalization ability of a classifier (i.e., its ability to distinguish between classes, such as the situation in which the exemplars were found) (Tajiri et al. 2010). We extracted mel frequency cepstral coefficients (MFCCs), representing the envelope of the short-time power spectrum as determined by the shape of the vocal tract (Logan 2000; Mielke and Zuberbühler 2013; Fedurek et al. 2016). The basic idea behind the extraction of MFCCs is to obtain a comprehensive representation of the frequencies that compose an audio excerpt, while emphasising certain frequency bands. Larger filters are used to extract acoustic features in higher frequencies, and the Mel-scale-related function enlarges filters as frequencies get higher. Additionally, the cepstral representation enables extracting precise information that is less dependent from the characteristics of the vocalizer.

We subdivided acoustic exemplars into windows of 25ms segments, with 10ms steps between two successive segments, to account for signal changes in time. We warped 26 spectral bands and returned 13 cepstra, resulting in feature dimensions of 13 values each. We then calculated the mean and co-variances of each cepstrum over the collection of feature segments, resulting in a 13-value vector and a 13 x 13-value matrix, concatenated to 104-unit vectors. We also applied feature scaling to values between 0 and 1.

Feature selection

Prior to classification, we conducted a feature selection procedure by reducing the number of features to a set of reliable features to explain the maximum variance in a given data set. We applied a *t*-test on each feature dimension by comparing values of the feature dimensions that were sorted by predefined class labels (e.g., situation ‘Threat’ vs. situation ‘No threat’). For each comparison, we used randomly determined 75% of the samples and re-ran the procedure five times. MATLAB's MathWorks webpage provides a tutorial on feature selection and all associated functions. Such feature selection procedure is called a filter approach, where general characteristics are evaluated for the selection without subjecting the dimensions to a classifier. The feature variance and feature relevance (i.e., additional improvement for each added feature) determine the importance of features. Since this is done in the pre-processing steps prior to classification, it is a procedure that is uncorrelated to the classification algorithm. In our case, we chose the feature dimensions that resulted in the highest *t*-values.

Classification

We implemented support vector machines (SVMs) using the LIBSVM toolbox (Chang and Lin 2011). Classification consisted of training and testing phases: first, 80% of the exemplars were selected to constitute a training dataset. Those exemplars were given attributes that indicated their 'class' (i.e., exemplars are marked with an attribute that tells whether the exemplar was taken, for example, from situation 'Threat' = 1 or situation 'No threat' = 0) and the model was trained to separate optimally between classes. Next, the trained model was tested on the remaining 20% of the exemplars (i.e., the test dataset) for which the attribute was unknown to the model. The performance of the model (i.e., its capacity to assign the correct class or attribute to this exemplar of the test dataset) was then evaluated and compared to a baseline level (the proportion of exemplars pertaining to a particular class in the whole set of exemplars). Training and testing were always done on two classes.

We used a radial basis function (RBF) kernel and 5-fold cross-validated the parameters C and Gamma with separate smaller data sets (for details see Fedurek et al. 2016). We used the top 40 feature dimensions for classification (see 'feature selection' above), omitting all others. We obtained performance scores from the models that were trained and tested on the same labels (basic classification procedure) and from cross-comparisons of conditions, such as training on situation 'Threat' and testing on one type of maternal behaviour. Previous work has shown that this method provides useful insights into the nature of information coding (Caldara and Abdi 2006; Fedurek et al. 2016; Dahl et al. 2018).

We compared the scores of correct classifications (i.e., the proportion of exemplars being assigned their true attribute [or 'class']) to a baseline level (i.e., the actual proportion of exemplars that correspond to a given attribute or class in the set of exemplars) using one-tailed two-sample *t*-tests. Two methods (Holm and Benjamini-Hochberg) were used to correct for multiple comparisons (Holm 1979; Benjamini and Hochberg 1995). To ensure that no single individual unduly influenced the outcome of the classification, a leave-one-out method was used, in which the general classification procedure was re-run eight times by omitting exemplars attributed to one infant in each run. We tested for a significant interaction between leave-one-out runs and the individual comparisons using a two-way ANOVA test.

Feature evaluation

In order to evaluate the features with highest contribution to the classification of certain attributes, we determined the extent to which comparisons shared similar feature dimensions for situations ('No threat' vs. 'Threat'; 'No threat' vs. 'Separation'; 'Threat' vs. 'Separation'). We additionally determined the top 10 feature dimensions of each comparison, as outlined above, and correlated the feature numbers of those feature dimensions in a pairwise fashion using Spearman's rank correlation tests. Feature numbers refer to the topological organization of the mel-frequency space, such that close feature numbers are indicative of similar underlying structures accounting for both comparisons. To determine the feature dimensions that are critical for the classification of exemplars, we assessed whether feature dimensions have been repeatedly used by the classifier overall in the classification. We therefore considered the 15 types of comparisons regarding infants' problems (e.g., 'Conflict' vs 'No reason'). As for the distance between the infant and the mother, we considered the six types of comparisons (e.g., 'Supported' vs. 'Contact'). Regarding maternal behaviour, we considered four yes-no comparisons, namely for ‘Gaze’, ‘Approach’, ‘Collection’ and ‘Vocalization’. We then examined the empirical distribution of the 40 feature dimensions used in the feature selection algorithms to determine the top ten contributors. Note that the choice of 10 features was arbitrary. As a baseline, a random distribution of “best features” for each comparison was determined by randomly selecting 10 out of 104 features. The frequency distribution across all comparisons was determined and 95% confidence intervals were calculated by running the procedure 1,000 times. We then reconstructed the underlying frequency bands of significant feature dimensions, resulting in feature maps. In a further step, we calculated differences in feature importance by pairwise contrasting feature maps. We contrasted two feature maps by subtracting corresponding frequency values, reflecting the occurrence of one particular mean cepstrum or co-variance of two cepstra, if at least one of the corresponding two values from the two feature maps was significant.

Results

Classification of exemplars across situations and problems

The SVM classifier revealed that the infants produced distress calls when separated from their mother (situation: ‘Separation’), when seemingly exposed to danger such as aggression or the experience of pain (situation: ‘Threat’) or for contexts we deemed non-threatening (situation: ‘No threat’). Classification accuracy for all 3 situations was high and significantly higher than baseline (Figure 1A; ESM Table 2).

The results of the leave-one-out procedure was not consistent with the effects observed due to single subjects, as indicated by an insignificant interaction between situations and the leave-one out runs ($F(14,720) = 1.36$, MS = 17.16, $p = 0.17$). This suggests our results are not unduly influenced by single individuals.

Can the acoustics of exemplars code for the exact problem the infants were dealing with? In some of the ‘Separation’ situations, physical distance was initiated by the mother and the infant was calling in response to the movement of the mother (problem: ‘Active separation’; ESM Table 1). In other episodes of the ‘Separation’ situations, the infant was already away from the mother and started calling in the absence of specific travelling movements from the mother (problem: ‘Passive separation’; ESM Table 1). Similarly, ‘Threat’ situations could involve different problems. In some cases of ‘Threat’, infants could be engaged in an activity or action that appeared to trigger physical pain such as due to the mother moving abruptly, or during rough social play (problem: ‘Pain’; ESM Table 1). In others, infants found themselves in a threatening social environment, such as when there was aggression in the vicinity (problem: ‘Social danger’; ESM Table 1).

Finally, the situation of ‘No threat’ could create conflicts of interests for the infants (problem: ‘Conflict’; ESM Table 1), such as when the mother and the infant appeared to disagree on travel decisions or food provisioning (e.g., mother refusing infant’s access to the nipple). In other instances of ‘No threat’, there was no obvious trigger that induced distress calls in the infant, and no specific reason could be identified (problem: ‘No reason’; ESM Table 1).

With the exception of ‘Active separation’ and ‘No reason’ problems, the classification accuracy for all other problems was significantly higher than baseline (Figure 1B and ESM Table 3). Interestingly, the discrimination of problems that belonged to the same situation (e.g., problems ‘Pain’ and ‘Social danger’, both of which occurred in the situation ‘Threat’) lead to less accurate classification, despite being significantly different from the baseline (ESM Table 3). The classifier’s discrimination of the problems was most accurate when contrasting the problems that occurred in the ‘Threat’ situation to problems that occurred in the ‘No threat’ situation. Similarly, contrasts, in which one of the problems occurred in the ‘Threat’ situation and the other in the ‘Separation’ situation yielded high classification scores. Classification of contrasts involving problems in the 'Separation' vs. 'No threat' situations appeared to be lower than that comparing problems in the 'Threat' vs. other situations ('Threat' vs. 'No threat': $M_1 = 85.30$, $SD_1 = 6.14$; $M_2 = 68.46$, $SD_2 = 8.25$; $t(78) = 14.13$, $p < 0.001$; 'Threat' vs 'Separation': $M_1 = 87.74$, $SD_1 = 8.39$; $M_2 = 68.46$, $SD_2 = 8.25$; $t(78) = 16.19$, $p < 0.001$). As shown in ESM Figure 4A, age did not appear to be driving the observed pattern, as the majority of exemplars were found around the same ages.

Feature analysis revealed similar rankings for contrasts of 'Threat' vs. 'No threat' and 'Threat' vs. 'Separation'. The correlation of feature numbers at a given rank in both contrasts yielded a positive correlation ($r_s = .93$, n = 10, $p < .001$, Figure 1C). This was not the case for pairs of contrasts involving either 'Separation' ($r_s = .47$, n = 10, $p = .18$) or 'No threat' ($r_s = .61$, n = 10, $p = .07$; Figure 1C). This indicates that the classification between pairs of situations involving 'Threat' relied on similar feature dimensions.

We also examined which feature dimensions most strongly accounted for the classification of exemplars according to the specific problems the infants were exposed to. This analysis revealed key contributions from the first, third and fourth mean cepstra, accompanied by covariances of cepstra corresponding with frequency bands around 342.59 to 1074.07 Hz and 50 to 781.48 Hz.

Classification of exemplars across distances between the infant and the mother

Our model classified distances at levels higher than baseline (Figure 2A and ESM Table 4). The leave-one-out procedure revealed no interaction between distances and leave-one-out runs ($F(5,288) = 0.92$, MS = 17.44, $p = 0.47$).

Further, we contrasted distance classes by assigning categorical units of distance (1 = most proximal ['Supported'] and 4 = most distal ['Beyond']) and by computing the relative distance between classes. We found that classification accuracy significantly increased as relative distance increased ($F(2,59) = 35.12$, MS = 16.21, $p < .001$; Figure 2B). This suggests that variation in classification performance reflects coding of distance in the exemplars. The acoustics of calls produced when the relative distance between the mother and the infant was more similar at episode onset were harder to discriminate than those produced when the mother and the infant were relatively more distant to each other.

Examination of feature dimensions that most contribute to the classification of distances revealed that features that are of significance for two or more comparisons were significantly different from the baseline. Indeed, the empirical number count 2 significantly surpassed a random distribution of number counts, relative to the total number of comparison (here 6) (Figure 2C). The third and fifth mean cepstra, accompanied by covariances of cepstra corresponding with frequency bands in the lower to mid-range of the Hz-spectrum, contributed most to classification (Figure 2D).

Classification of exemplars and maternal behaviour

Our model discriminated the probability of the mother to gaze towards, approach, collect the infant and vocalize contingently to the infant's distress calls accurately and above baseline (Figure 3A; ESM Table 5). Although discrimination accuracy was always above baseline level, it was lower when predicting maternal collection and vocalization behaviours (Figure 3A). The leave-one-out procedure revealed no significant interaction between individual runs (N = 8) and maternal behaviour ($F(3,192) = 0.67$, MS = 1.87, $p = 0.570$).

To evaluate the features most contributing to determining whether the mother did or did not gaze, approach, collected the infant or vocalized contingently to the infant's calling, we plotted the mean expression of each feature dimension for both the presence ('Yes') and the absence ('No') of a given maternal behaviour. Selected features were often located at the outer edge of the distribution (ESM Figure 2B-E), suggesting that they account for one class significantly stronger than for the other class. The contrasts that yielded high performance scores showed more distinct feature separation (e.g., 'Gaze' plot - ESM Figure 2B) than the contrasts that yielded lower classification (e.g., 'Collection' plot - ESM Figure 2D). We also examined which feature dimensions accounted for the classification of exemplars the most according to maternal behaviour. This analysis revealed that features that are of significance for two or more comparisons were significantly different from the baseline. The empirical number count 2 significantly surpassed a random distribution of number counts, relative to the total number of comparisons (here 4; Figure 3B). Notably, the first, third, fourth and the 12th mean cepstra, accompanied by accompanied by covariances of cepstra corresponding with frequency bands around 342.59 to 1074.07 Hz and 50 to 781.48 Hz, contributed significantly to the classification (Figure 3C). The feature distribution of maternal behaviour classification resembles the feature distribution of problems faced by the infants (Figure 1E).

Mother-infant interactions

To further understand how information-coding of the situation in the exemplars is associated with a particular configuration of maternal behaviour, we trained a model on discriminating situations and infants' problems, and tested its performance in classifying the presence of the various types of maternal behaviour.

Our model's discrimination accuracy was significantly higher than baseline for contrasts of the 'Threat' and 'No threat' situations across all maternal behaviour types (M_{emp} = 69.38, SD_{emp} = 9.39; M_{base} = 54.65, SD_{base} = 2.87; $t(78)$ = 9.48, $p < 0.001$), as well as of the 'Threat' and 'Separation' situations (M_{emp} = 72.31, SD_{emp} = 8.87; M_{base} = 55.33, SD_{base} = 2.67; $t(78)$ = 11.57, $p < 0.001$; see Figure 4A). However, contrasts of the other two other situations ('No threat' and 'Separation') were not successfully discriminated based on maternal behaviour (M_{emp} = 55.33, SD_{emp} = 4.33; M_{base} = 54.90, SD_{base} = 2.87; $t(78)$ = 0.52, p = 0.300).

Testing the model with problems, we found all comparisons to be significant ($p < .05$; ESM Table 6) except for the contrasts of the 'No reason' vs. 'Active separation' problems, and of the 'Conflict' vs. 'No reason' problems (Figure 4B). Gazing was a relatively reliable action of the mother, particularly during problems associated with Situation ‘Threat’. On the other hand, ‘Collection’ and ‘Vocalization’ were found to be more modest in discriminating between problems.

To address the extent to which structural differences of exemplars reflect structural differences encoding distance and predicting maternal behaviour, we conducted pairwise contrasts of the feature maps for situation, distance and maternal behaviour. We found that, when comparing situation and distance, certain features were selective for situation that related strongly to the first frequency band ranging from 50 to 342.59 Hz and co-varied with frequency bands ranging from 488.89 to 927.78 Hz. In addition, the frequency band centring at 1366.70 Hz co-varied with the second frequency band ranging from 196.30 to 488.89 Hz and the fifth ranging from 635.19 to 927.78 Hz (Figure 4C). On the other hand, distance was selectively coded in the second frequency band ranging from 196.30 to 488.89 Hz, co-varying with multiple feature dimensions (Figure 4C). When comparing situation and maternal behaviour (Figure 4D), we found similar components involved, however, to a marginal degree, indicating that there are similar main features accounting for most of the classification of exemplars for both situation and maternal behaviour. Comparison between distance and maternal behaviour revealed similar results to the comparison between situation and distance (Figure 4E).

Discussion

Distress calls have been described as acoustically continuous or graded in various mammals, but studies (notably in humans) have raised the possibility that these calls may convey discrete information (Müller et al. 1974; Wiesenfeld et al. 1981; Brennan and Kirkland 1982; Fuller 1991; Soltis 2004). Here, we used a supervised machine learning approach to investigate distress calls of wild infant chimpanzees and evaluate whether they carry information about the nature of the external events experienced by the caller and whether they can be associated with particular maternal behaviours, which the calls may elicit.

We extracted exemplars from distress call episodes and used machine learning classification techniques. We found that a model trained on discriminating exemplars between a threatening situation ('Threat') from others was better at predicting the mother's behaviour, suggesting that distress calls during threatening situations rely on specific acoustic features. This is further exemplified by the fact that discriminations between exemplars found in situations 'Threat' vs. 'Separation' on the one hand, and 'Threat vs. No threat' on the other hand, relied on common features. Our results suggest that the acoustic information present in distress call episodes contain information about the type of external events triggering calling as well as the nature of the problem. Those results do not appear to be driven by the inclusion of particular individuals in the dataset (as indicated by insignificant leave-one-out procedures) or by the prominence of particular age classes for specific problems.

We also found that the distance between infants and their mothers at the onset of distress call episodes is associated with the particular structure of the exemplars, with exemplars being acoustically most distinctive as the distance between the infants and mothers increases. Another relevant finding was that whether the mothers would gaze at, approach, collect or vocalize at their infant during the call episode could be predicted by the acoustic characteristics of the exemplars. Further, we found that feature maps supporting the classifications of situations and maternal behaviour were relatively similar.

Our results suggest that, in principle, chimpanzee mothers could rely on the acoustic information contained in distress calls to make intervention decisions. Such decisions may be based on acoustic cues that are linked to the affective state of the infant and that reflect the intensity and severity of the problem (Weary et al. 1996; Briefer 2012). The ability to process these acoustic cues and react accordingly can enhance the fitness of both the infants vocally conveying their specific needs and the fitness of their close relatives, namely their mothers. This potentially fitness enhancing ability can be particularly important in a species in which infanticide by both males and females is commonly reported (Arcadi and Wrangham 1999; Watts and Mitani 2000; Townsend et al. 2007; Lowe et al. 2019, 2020).

However, it is not clear that mothers solely rely on acoustical cues. A number of other cues (infants' facial features, movements and gestures; reaction of other conspecifics; knowledge of past problems experienced by the infant) must also be playing a role in shaping mothers' reactions. Our findings only suggest that distress calls encode sufficient information to shape maternal reaction decisions, not that mothers only take acoustic cues into account. Moreover, we cannot argue that infant distress calls are the causes of the maternal behaviours examined here. It is possible that the mothers were to display the kind of protective behaviour they showed regardless of whether the infant called or not (because they were to do it anyway, or simply because they perceived the problems infants were then facing through other means). The current study does not allow us to disentangle whether, when the mothers responded to their infants' calling, they did it in response to the calls alone or they relied on numerous other cues. A potential way of examining this is by recording distress calls in situations where infants are clearly out of sight of the mothers, a scenario unlikely to happen with young chimpanzee infants, given their relative dependency to their mothers (Plooij et al. 1984).

One strength of the current study is its use of robust machine learning approaches to examine highly complex acoustic phenomena, which would otherwise be intractable. Besides the technical achievement the development of machine learning has undeniably brought in a number of fields of research (Riecken 2000; Sebe et al. 2005; Olsson 2009; Deng and Li 2013; Kelleher et al. 2015; Libbrecht and Noble 2015; Nithya and Ilango 2017), its value in the study of animal cognition is becoming clearer (Gerencsér et al. 2013; Wiltschko et al. 2015; Dahl et al. 2018). In a recent study, Fedurek and colleagues (Fedurek et al. 2016) were able to show that different phases of the complex pant-hoot calls of male chimpanzees carry information about particular features of the caller, notably its age and identity, revealing unprecedented details about how information can be acoustically represented.

A potential weakness of machine learning approaches is their technical difficulty. This is particularly true when used in combination with other technical approaches (such as the extraction of MFCCs) that are less directly amenable to biological interpretations than conventional acoustic metrics such as the description of certain frequency parameters, or the measurement of the fundamental frequency (Briefer 2012). In this study, we also used an automated routine for the extraction of exemplars representing the distress call episodes, which further complexified the approach.

Our results suggest that the distress calls of infant chimpanzees may be acoustically rich enough to convey information about the external events triggering the calls, and as such, may help caregivers to make important intervention decisions. Undoubtedly, infant distress calls are interpreted by the recipients in combination with other signals such as gestures, postures and other conspecifics' reactions that often accompany them (Hobaiter and Byrne 2011, 2014; Fröhlich et al. 2016, 2019; Fröhlich and Hobaiter 2018). Mothers also probably make inferences about the needs of the infants from past knowledge of problems experienced by the infant. How the recipients consider and integrate across these different sources of information needs to be addressed by future research.

References

Altmann J (1974) Observational study of behavior: sampling methods. Behaviour 49:227–266

Arcadi AC, Wrangham RW (1999) Infanticide in chimpanzees: review of cases and a new within-group observation from the Kanyawara study group in Kibale National Park. Primates 40:337–351

Barajas-Montiel SE, Reyes-Garcia CA (2006) Fuzzy support vector machines for automatic infant cry recognition. In: Intelligent Computing in Signal Processing and Pattern Recognition. Springer, pp 876–881

Bard KA (2000) Crying in infant primates: Insights into the development of crying in chimpanzees. In: Crying as a Sign, a Symptom, and a Signal: Clinical, Emotional and Developmental Aspects of Infant and Toddler Crying. Mac Keith Press, London, p 157

Bayart F, Hayashi KT, Faull KF, et al (1990) Influence of maternal proximity on behavioral and physiological responses to separation in infant rhesus monkeys (Macaca mulatta). Behavioral neuroscience 104:98

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57:289–300

Boll S (1979) Suppression of acoustic noise in speech using spectral subtraction. IEEE Transactions on acoustics, speech, and signal processing 27:113–120

Brennan M, Kirkland J (1982) Classification of infant cries using descriptive scales. Infant Behavior and Development 5:341–346

Briefer EF (2012) Vocal expression of emotions in mammals: mechanisms of production and evidence. Journal of Zoology 1–20. https://doi.org/10.1111/j.1469-7998.2012.00920.x@10.1111/(ISSN)1469-7998.SoundProduction

Caldara R, Abdi H (2006) Simulating the ‘other-race’effect with autoassociative neural networks: further evidence in favor of the face-space model. Perception 35:659–670

Chang C-C, Lin C-J (2011) LIBSVM: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST) 2:1–27

Chang C-Y, Li J-J (2016) Application of deep learning for recognizing infant cries. In: 2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW). IEEE, pp 1–2

Choqueuse V (2020) Vuvuzela sound denoising algorithm. Version 1.2.0.0URL https://fr.mathworks.com/matlabcentral/fileexchange/27912-vuvuzela-sound-denoising-algorithm?focused=b421486c-db15-d58a-6f4d-5026e5f66319&tab=function

Christensson K, Cabrera T, Christensson E, et al (1995) Separation distress call in the human neonate in the absence of maternal body contact. Acta paediatrica 84:468–473

Crockford C, Gruber T, Zuberbühler K (2018) Chimpanzee quiet hoo variants differ according to context. Royal Society open science 5:172066

Crockford C, Wittig RM, Mundry R, Zuberbühler K (2012) Wild chimpanzees inform ignorant group members of danger. Current Biology 22:142–146. https://doi.org/10.1016/j.cub.2011.11.053

Crockford C, Wittig RM, Zuberbühler K (2017) Vocalizing in chimpanzees is influenced by social-cognitive processes. Sci Adv 3:e1701742. https://doi.org/10.1126/sciadv.1701742

Dahl CD, Wyss C, Zuberbühler K, Bachmann I (2018) Social information in equine movement gestalts. Animal cognition 21:583–594

Deng L, Li X (2013) Machine learning paradigms for speech recognition: An overview. IEEE Transactions on Audio, Speech, and Language Processing 21:1060–1089

Dezecache G, Crockford C, Zuberbühler K (2019) The development of communication in alarm contexts in wild chimpanzees. Behav Ecol Sociobiol 73:104. https://doi.org/10.1007/s00265-019-2716-6

Eggeling WJ (1947) Observations on the ecology of the Budongo rain forest, Uganda. The Journal of Ecology 20–87

Ephraim Y, Malah D (1984) Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Transactions on acoustics, speech, and signal processing 32:1109–1121

Fedurek P, Zuberbühler K, Dahl CD (2016) Sequential information in a great ape utterance. Scientific reports 6:38226

Fischer J (1998) Barbary macaques categorize shrill barks into two call types. Animal behaviour 55:799–807

Fröhlich M, Hobaiter C (2018) The development of gestural communication in great apes. Behavioral Ecology and Sociobiology 72:194

Fröhlich M, Wittig RM, Pika S (2016) Should I stay or should I go? Initiation of joint travel in mother–infant dyads of two chimpanzee communities in the wild. Animal Cognition 19:483–500

Fröhlich M, Wittig RM, Pika S (2019) The ontogeny of intentional communication in chimpanzees in the wild. Developmental science 22:e12716

Fuller BF (1991) Acoustic discrimination of three types of infant cries. Nursing Research 40:156–160

Gerencsér L, Vásárhelyi G, Nagy M, et al (2013) Identification of behaviour in freely moving dogs (Canis familiaris) using inertial sensors. PloS one 8:e77814

Gilbert HR, Robb MP (1996) Vocal fundamental frequency characteristics of infant hunger cries: birth to 12 months. International journal of pediatric otorhinolaryngology 34:237–243

Goodall J (1986) The chimpanzees of Gombe: Patterns of behavior. Harvard University Press

Green PA, Brandley NC, Nowicki S (2020) Categorical perception in animal communication and decision-making. Behav Ecol. https://doi.org/10.1093/beheco/araa004

Hobaiter C, Byrne RW (2011) The gestural repertoire of the wild chimpanzee. Animal cognition 14:745–767

Hobaiter C, Byrne RW (2014) The meanings of chimpanzee gestures. Current Biology 24:1596–1600

Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics 65–70

Illingworth RS (1955) Crying in Infants and Children. Br Med J 1:75–78

Kelleher JD, Mac Namee B, D’arcy A (2015) Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT press

Levréro F, Mathevon N (2013) Vocal signature in wild infant chimpanzees. American Journal of Primatology 75:324–332

Libbrecht MW, Noble WS (2015) Machine learning applications in genetics and genomics. Nature Reviews Genetics 16:321–332

Lingle S, Riede T (2014) Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. The American Naturalist 184:510–522

Lingle S, Wyman MT, Kotrba R, et al (2012) What makes a cry a cry? A review of infant distress vocalizations. Current Zoology 58:698–726

Logan B (2000) Mel frequency cepstral coefficients for music modeling. In: Ismir. pp 1–11

Lowe AE, Hobaiter C, Asiimwe C, et al (2020) Intra-community infanticide in wild, eastern chimpanzees: a 24-year review. Primates 61:69–82

Lowe AE, Hobaiter C, Newton-Fisher NE (2019) Countering infanticide: chimpanzee mothers are sensitive to the relative risks posed by males on differing rank trajectories. American journal of physical anthropology 168:3–9

MacLean PD (1985) Brain evolution relating to family, play, and the separation call. Archives of general psychiatry 42:405–417

May B, Moody DB, Stebbins WC (1989) Categorical perception of conspecific communication sounds by Japanese macaques, Macaca fuscata. The Journal of the Acoustical Society of America 85:837–847. https://doi.org/10.1121/1.397555

Mielke A, Zuberbühler K (2013) A method for automated individual, species and call type recognition in free-ranging animals. Animal Behaviour 86:475–482

Mohri M, Rostamizadeh A, Talwalkar A (2018) Foundations of machine learning. MIT press

Müller E, Hollien H, Murry T (1974) Perceptual responses to infant crying: Identification of cry types. Journal of Child Language 1:89–95

Newman JD (2007) Neural circuits underlying crying and cry responding in mammals. Behavioural brain research 182:155–165

Nithya B, Ilango V (2017) Predictive analytics in health care using machine learning tools and techniques. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, pp 492–499

Olsson F (2009) A literature survey of active machine learning in the context of natural language processing

Plooij FX, Bygott D, Sackett GP (1984) The behavioral development of free-living chimpanzee babies and infants. Ablex Norwood, NJ

Porter FL, Miller RH, Marshall RE (1986) Neonatal pain cries: effect of circumcision on acoustic features and perceived urgency. Child development 790–802

Reynolds V (2005) The chimpanzees of the Budongo forest: Ecology, behaviour and conservation. OUP Oxford

Riecken D (2000) Introduction: personalized views of personalization. Communications of the ACM 43:26–28

Saraswathy J, Hariharan M, Yaacob S, Khairunizam W (2012) Automatic classification of infant cry: A review. In: 2012 International Conference on Biomedical Engineering (ICoBE). pp 543–548

Schel AM, Townsend SW, Machanda Z, et al (2013) Chimpanzee alarm call production meets key criteria for intentionality. PloS one 8:e76674. https://doi.org/10.1371/journal.pone.0076674

Sebe N, Cohen I, Garg A, Huang TS (2005) Machine learning in computer vision. Springer Science & Business Media

Soltis J (2004) The signal functions of early infant crying. Behavioral and brain sciences 27:443–458

Tajiri Y, Yabuwaki R, Kitamura T, Abe S (2010) Feature extraction using support vector machines. In: International Conference on Neural Information Processing. Springer, pp 108–115

Townsend SW, Slocombe KE, Thompson ME, Zuberbühler K (2007) Female-led infanticide in wild chimpanzees. Current Biology 17:R355–R356

Turesson HK, Ribeiro S, Pereira DR, et al (2016) Machine learning algorithms for automatic classification of marmoset vocalizations. PloS one 11:e0163041

Versteegh M, Kuhn J, Synnaeve G, et al (2016) Classification and automatic transcription of primate calls. The Journal of the Acoustical Society of America 140:EL26–EL30

Wasz-Höckert O, Partanen TJ, Vuorenkoski V, et al (1964) The identification of some specific meanings in infant vocalization. Experientia 20:154–154

Watts DP, Mitani JC (2000) Infanticide and cannibalism by male chimpanzees at Ngogo, Kibale National Park, Uganda. Primates 41:357–365

Weary DM, Lawson GL, Thompson BK (1996) Sows show stronger responses to isolation calls of piglets associated with greater levels of piglet need. Animal Behaviour 52:1247–1253. https://doi.org/10.1006/anbe.1996.0272

Wiener SG, Bayart F, Faull KF, Levine S (1990) Behavioral and physiological responses to maternal separation in squirrel monkeys (Saimiri sciureus). Behavioral neuroscience 104:108

Wiesenfeld AR, Malatesta CZ, Deloach LL (1981) Differential parental response to familiar and unfamiliar infant distress signals. Infant Behavior and Development 4:281–295

Wiltschko AB, Johnson MJ, Iurilli G, et al (2015) Mapping sub-second structure in mouse behavior. Neuron 88:1121–1135

Zeskind PS, Sale J, Maio ML, et al (1985) Adult perceptions of pain and hunger cries: A synchrony of arousal. Child development 549–554

Tables

Table 1

List of infants, estimated birthdate, minimum and maximum age in months over the course of the study, sex, number of contributed distress call episodes used in acoustic extraction, and number of extracted exemplars.

ID	Birthdate	Min. age	Max. age	Sex	Call episodes (N)	Extracted exemplars (N)
KF	26/03/2014	0.13	24.26	M	9	37
OZ	16/09/2014	1.38	18.16	M	48	700
MZ	27/10/2015	4.20	5.28	M	32	147
RY	08/10/2013	4.75	7.74	M	9	22
HM	26/10/2013	4.98	6.82	F	5	11
KO	07/09/2014	6.10	19.57	M	21	74
KJ	07/07/2013	7.02	32.69	M	49	333
KV	26/11/2014	15.67	17.80	M	5	6

Figures

Figure 1. Model performance to classify across the situations and infants' problems, and evaluation of key features contributing to classification

(A, B) Accuracy values of classification runs are shown for all contrasts between situations (panel A) and problems (panel B). Blue horizontal bars indicate the means, black dots indicate the results of cross-validation and red horizontal lines indicate the minimal and maximal values of the leave-one-out procedure. (C) The feature sets accounting for individual contrasts the most are compared and correlated. Dots show individual samples; relationships between them are indicated by regression lines. (D) Feature evaluation procedure, showing the occurrence count (x-axis), reflecting the number of times a particular feature dimension was among the top 10 feature dimensions across all comparisons, and the relative frequency of n-counts (y-axis). The empirical distribution is shown in red, a random distribution (solid line) in black, and 95% confidence-interval in dotted lines. N-counts of 4 or more are significantly over-represented. (E) X- and Y-axes represent frequency bands. Significant feature dimensions are colour-coded and structurally aligned in a frequency-transformed representation. Blue dots indicate significant mean cepstra, red dots indicate positive co-variances of cepstra, and green dots indicate negative co-variances of cepstra. Grey dots indicate non-significant feature dimensions. The size of dots indicates their relative importance: the larger the dot the more frequently a feature dimension has been used across all comparisons.

Figure 2. Model performance to classify across distances between the mother and the infant, and evaluation of key contributing features

(A) Accuracy values of classification runs are shown, contrasting various categorical distances between the infant and the mother. Colour-code is the same as in Figure 1 panels A & B. (B) Accuracy scores are plotted according to the relative distances compared (max 3: Supported

[Supported] - Beyond arm's reach [Beyond]). Accuracy values are shown as z-scores. Red line represents the linear regression fit. (C) Feature evaluation procedure, showing the occurrence count (x-axis), reflecting the number of times a particular feature dimension was among the top 10 feature dimensions across all comparisons, and the relative frequency of n-counts (y-axis). Colour-code is the same as in Figure 1D. N-counts of 2 or more are significantly over-represented. (D) Significant feature dimensions (see Figure 1E for colour-codes).

Figure 3. Classification performance per type of maternal behaviour and evaluation of contributing features

(A) Accuracy values of classification runs for maternal behaviour. Colour-code is the same as in Figures 1 panels A & B. (B) Feature comparison, showing the occurrence count (x-axis), reflecting the number of times a particular feature dimension was among the top 10 feature dimensions across all comparisons, and the relative frequency of n-counts (y-axis). Colour-code is the same as in Figure 1D. N-counts of 2 or more are significantly over-represented. (C) Significant feature dimensions are colour-coded and structurally aligned in a frequency-transformed representation, following Figure 1E.

Figure 4. Mother-infant interaction classification models

(A & B) Accuracy values of classification runs for models trained on situation (A) or infants' problems (B) contrasts (y-axis) and tested on maternal behaviour contrasts (x-axis). The colour represents the level of accuracy (0% corresponds to the white colour; 100% to the red colour). (C, D, E) Feature map comparison. Feature maps were contrasted by subtracting corresponding frequency values, reflecting the occurrence of one particular mean cepstra or co-variance of two cepstra, if at least one of the corresponding two values from the two feature maps was significant. Green indicates greater importance for the first term (e.g., 'situation' in the extremeleft panel), red for the second term of the comparison (e.g., ‘maternal behaviour’ in the extreme-right panel).

Figure 1

Click here to access/download;Figure;figure1.tiff

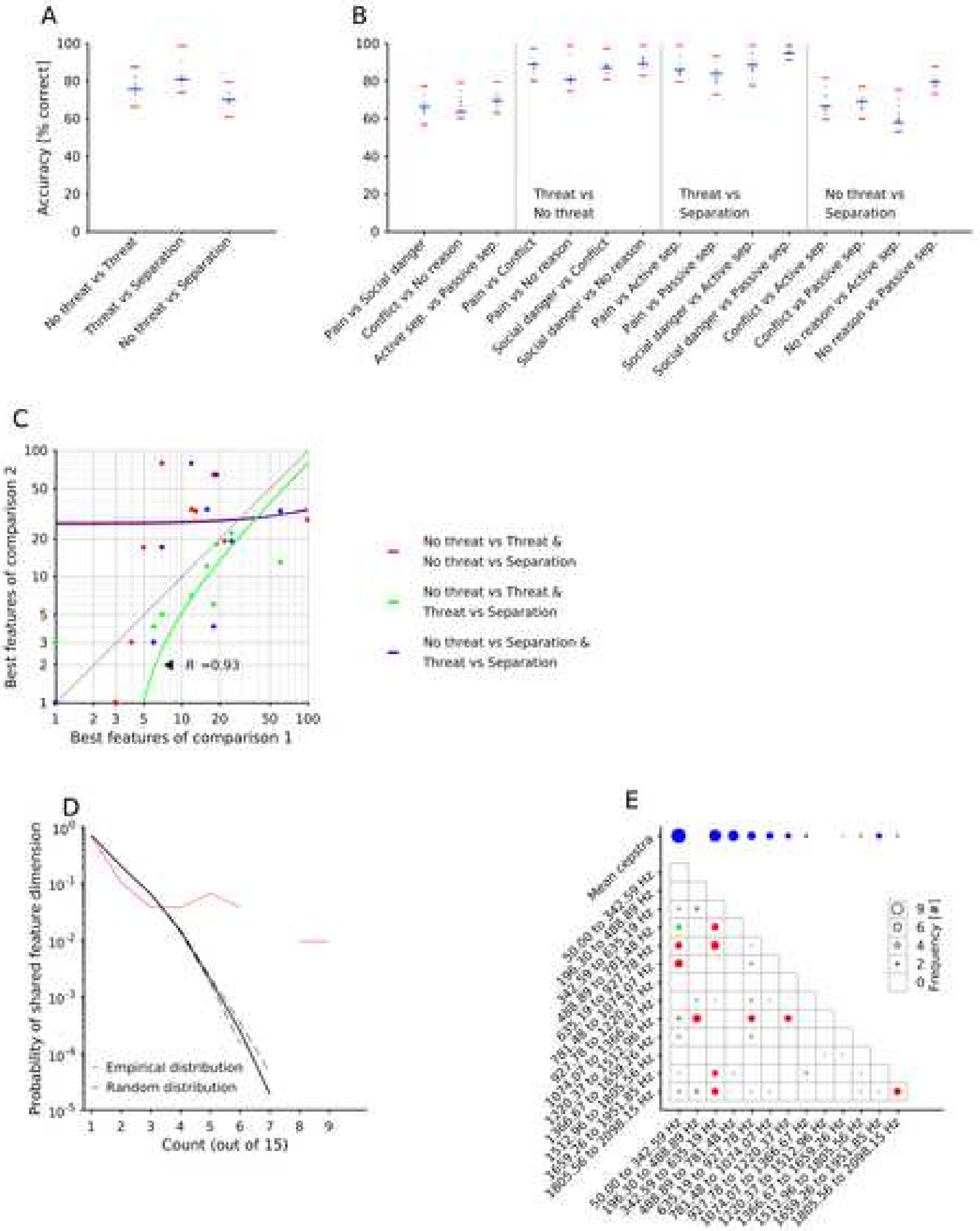


Figure 2

Click here to access/download;Figure;figure2.tiff

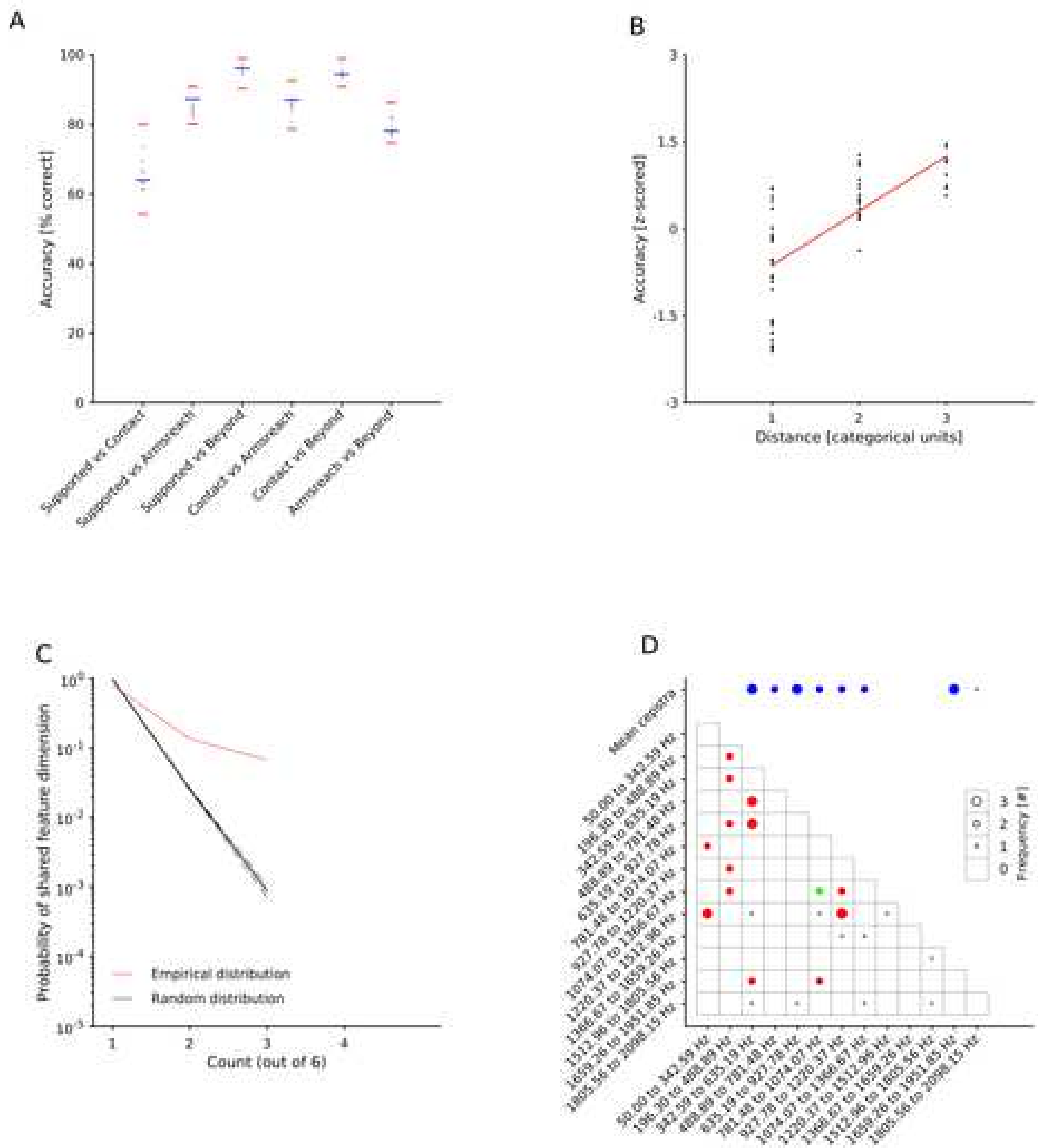


Figure 3

Click here to access/download;Figure;figure3.tiff

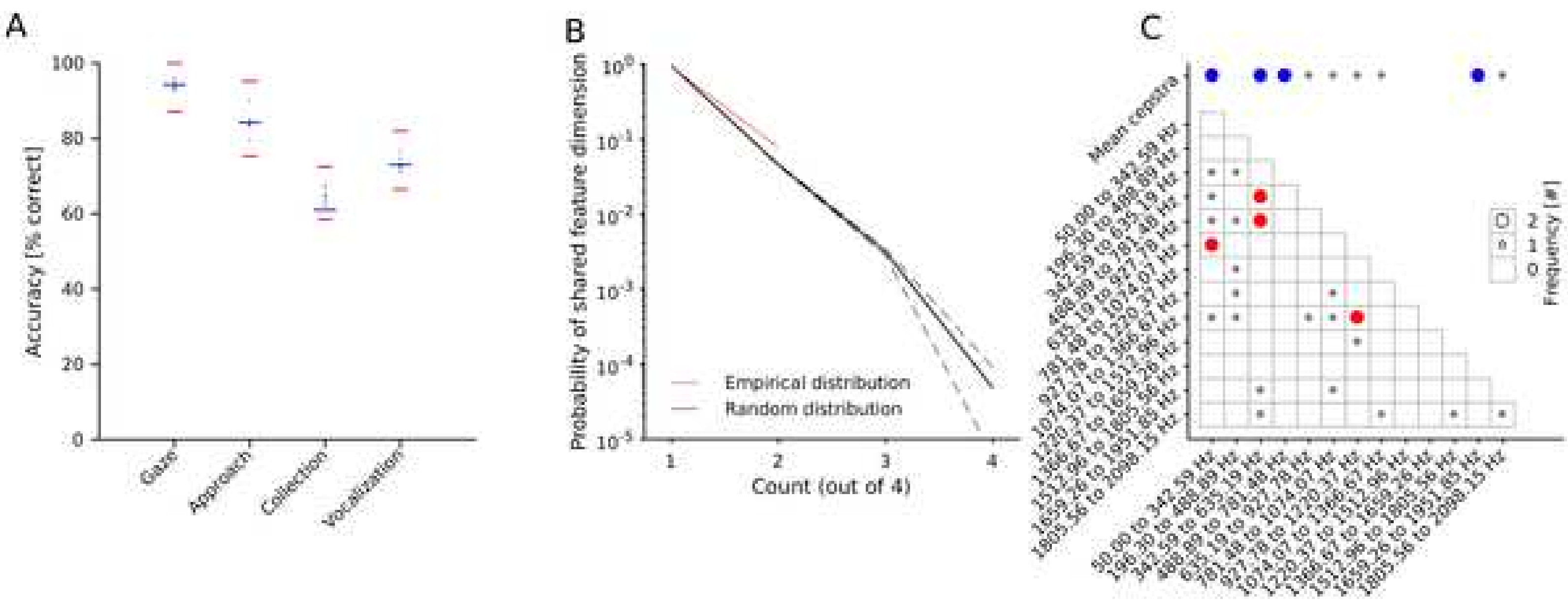
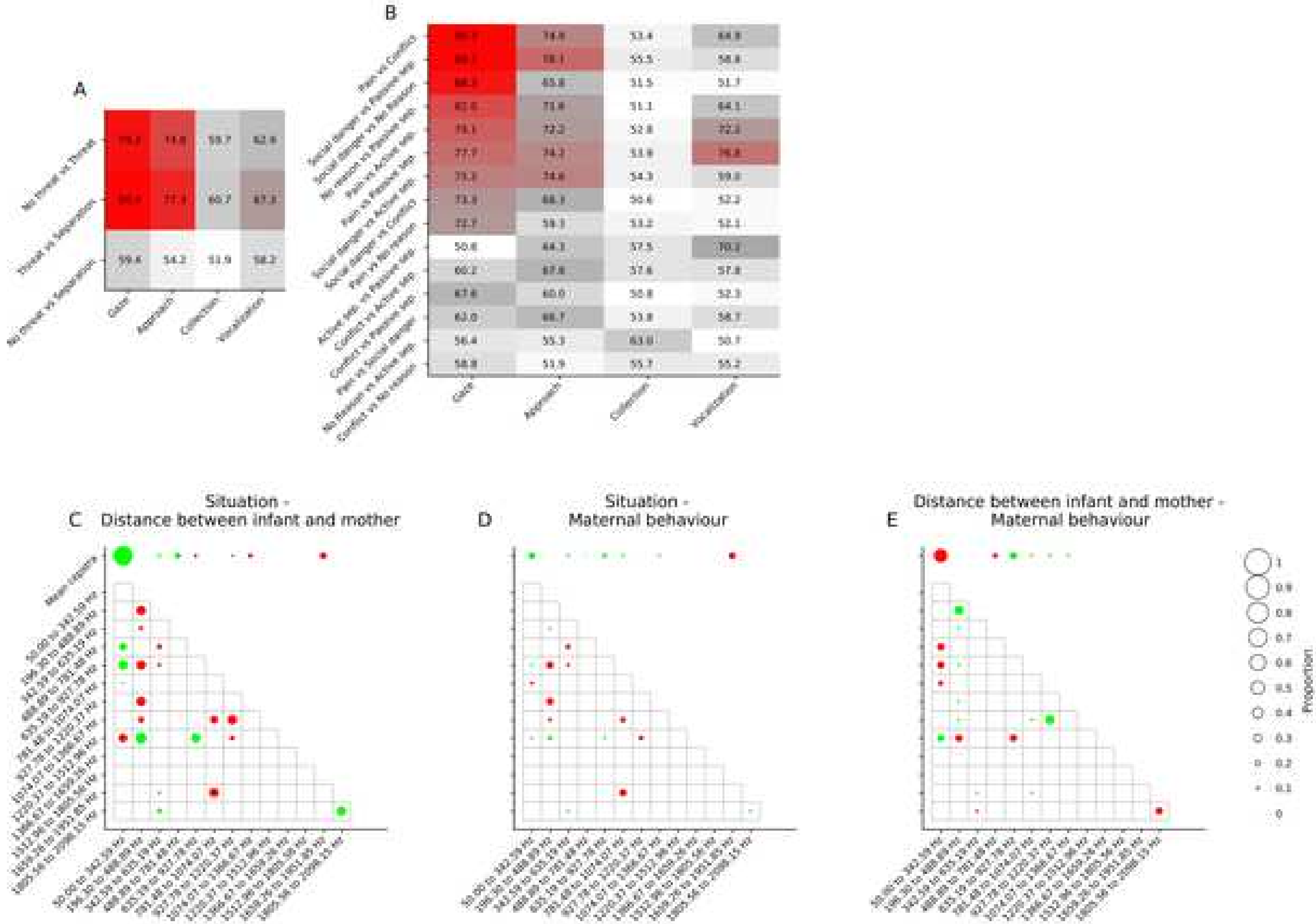


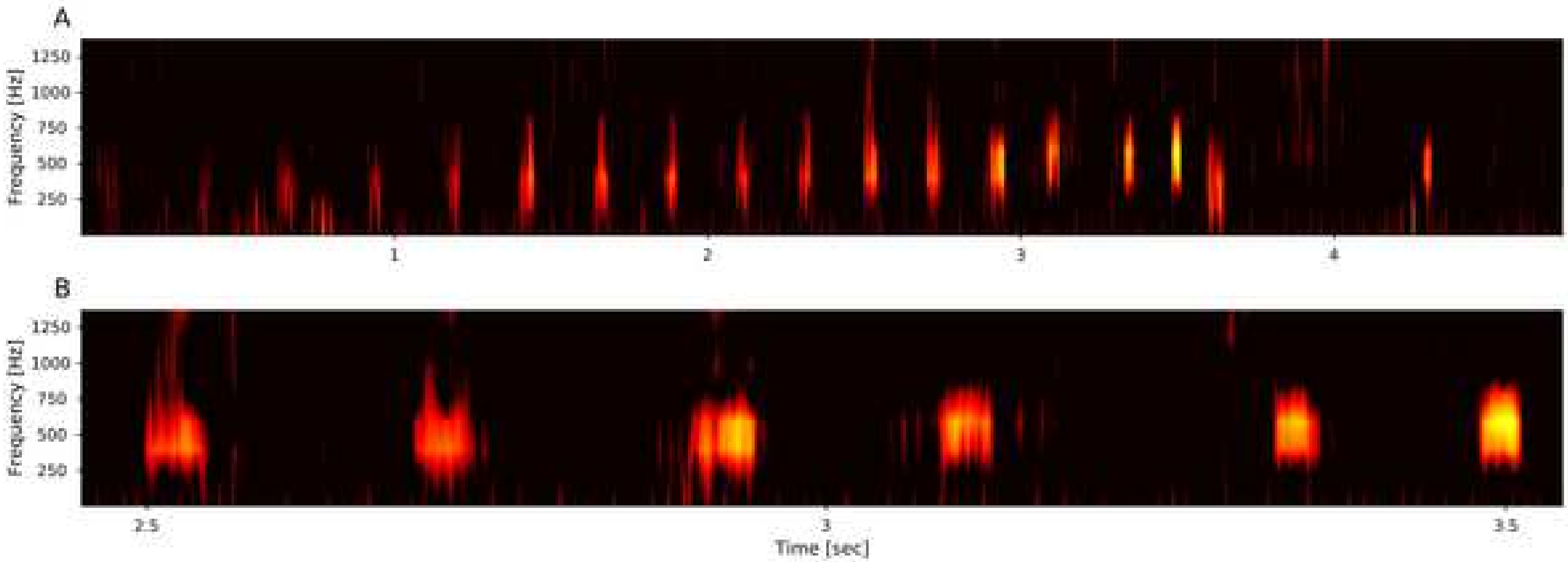
Figure 4

Click here to access/download;Figure;figure4.tiff



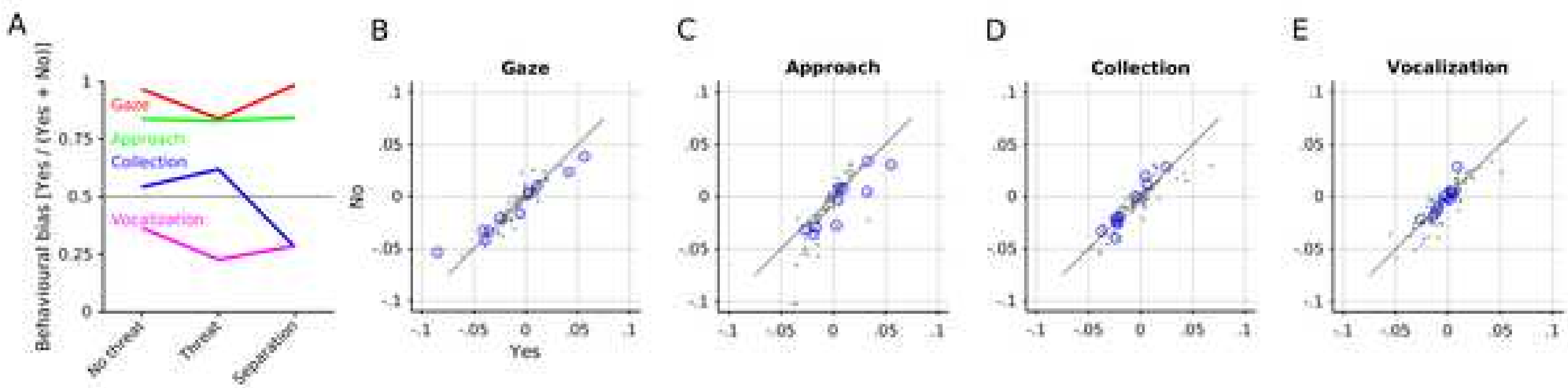
ESM Figure 1

Click here to access/download;Figure;ESMfig01.tiff



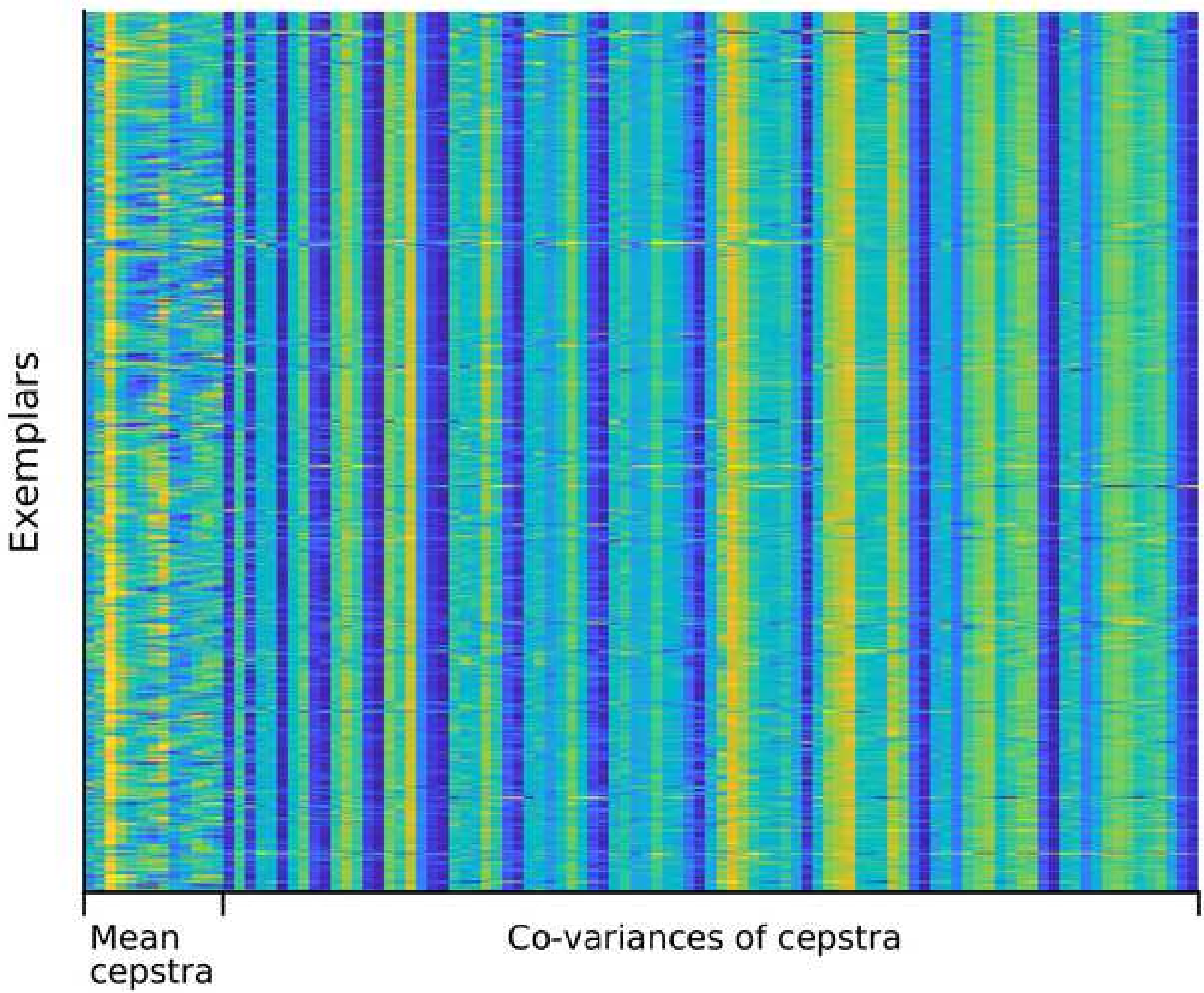
ESM Figure 2

Click here to access/download;Figure;ESMfig02.tiff



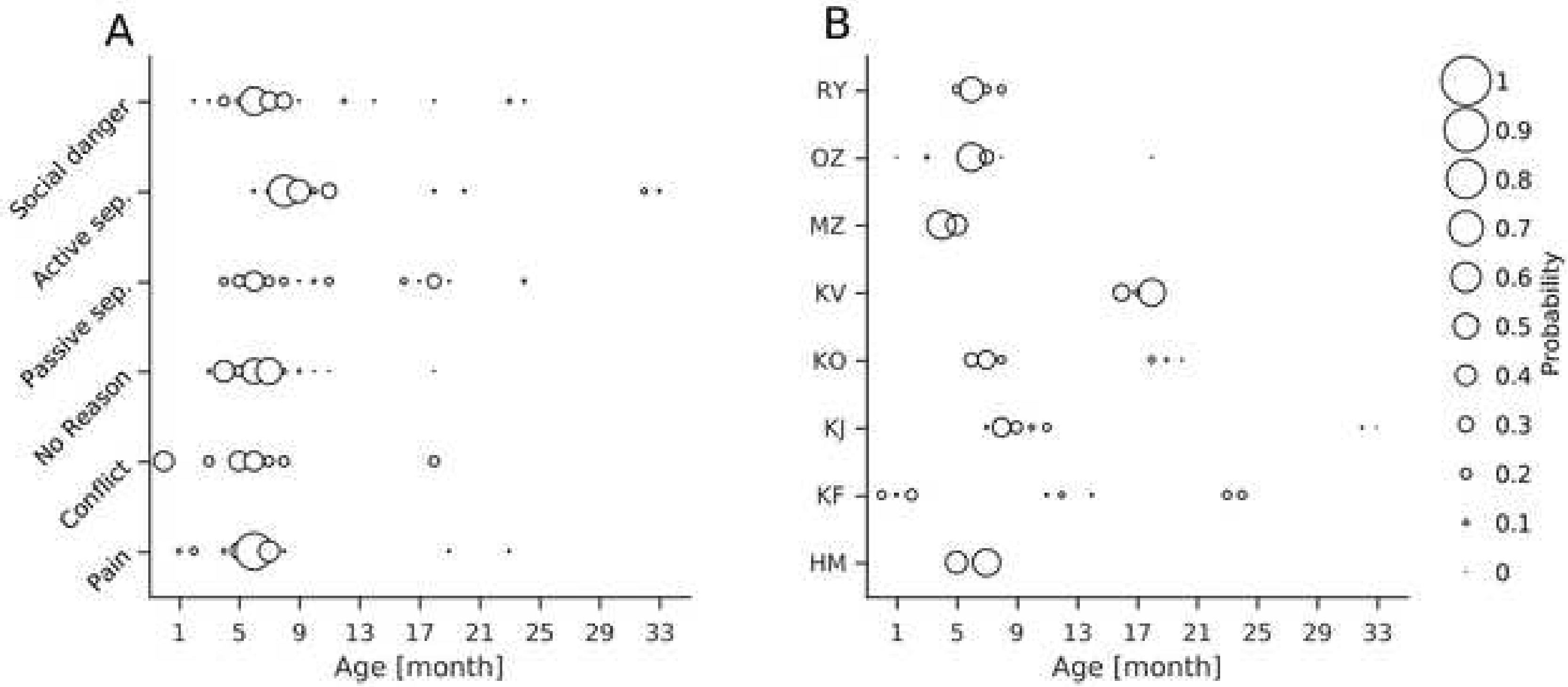
ESM Figure 3

Click here to access/download;Figure;ESMfig03.tiff



ESM Figure 4

Click here to access/download;Figure;ESMfig04.tiff



Click here to access/download
Supplementary Material
Electronic Supplementary Material.docx