



HAL
open science

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

Guillaume Dezecache, Klaus Zuberbühler, Marina Davila-Ross, Christoph Dahl

► **To cite this version:**

Guillaume Dezecache, Klaus Zuberbühler, Marina Davila-Ross, Christoph Dahl. A machine learning approach to infant distress calls and maternal behaviour of wild chimpanzees. *Animal Cognition*, 2021, 24 (443-455), 10.1007/s10071-020-01437-5 . hal-02983026

HAL Id: hal-02983026

<https://hal.science/hal-02983026>

Submitted on 4 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

[Click here to view linked References](#)

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

2 Guillaume Dezecache^{a,b,c,d,*}, Klaus Zuberbühler^{a,b,e}, Marina Davila-Ross^c & Christoph D.

3 Dahl^{a,f,g,*}

4 ^aInstitute of Biology, University of Neuchâtel, Neuchâtel, Switzerland;

5 ^bBudongo Conservation Field Station, Masindi, Uganda;

6 ^cDepartment of Psychology, University of Portsmouth, Portsmouth, England, United

7 Kingdom;

8 ^dUniversité Clermont Auvergne, CNRS, LAPSCO, Clermont-Ferrand, France;

9 ^eSchool of Psychology and Neuroscience, University of St Andrews, St Andrews, Scotland,

10 United Kingdom;

11 ^fGraduate Institute of Mind, Brain and Consciousness, Taipei Medical University, Taipei,

12 Taiwan;

13 ^gBrain and Consciousness Research Center, Taipei Medical University Shuang-Ho Hospital,

14 New Taipei City, Taiwan

15

16 *Correspondence:

17 Guillaume Dezecache <guillaume.dezecache@gmail.com> ; +33473406458; LAPSCO-UMR

18 -CNRS 6024, 17 Rue Paul Collomp, 63000 Clermont-Ferrand, France

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

20

21 **Acknowledgments**

22 We thank the Uganda Wildlife Authority (UWA) and the Uganda National Council for

23 Science and Technology (UNCST) for permission to conduct the study, and Geoffrey

24 Muhanguzi, Caroline Asiimwe, Sam Adué and Monday M'Botella for their support in the

25 field. We are grateful to two anonymous Reviewers, the Editor and Bahar Tuncgenc for

26 helpful comments on the manuscript.

27 **Abstract**

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

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

41

42 **Declarations**

43 **Funding**

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

50

51 ***Authors' contributions***

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

55

56 ***Conflict of interests***

57 None.

58

59 ***Code availability***

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

61 **Introduction**

62 Distress calls are the most primitive mammalian vocalizations (MacLean 1985; Newman
63 2007). They appear early in ontogeny (Illingworth 1955) and are highly preserved in
64 phylogeny, with a simple structure (tonal sound with a chevron or descending shape) also
65 present in species beyond the class Mammalia (Lingle et al. 2012). Their function is most
66 likely to recruit help (Soltis 2004; Lingle et al. 2012; Lingle and Riede 2014), but it is
67 uncertain whether caregivers can base their reactions and inferences about the nature of the
68 disturbance on the acoustics of the calls alone, or whether they must also rely on contextual
69 cues.

70

71 In humans, at least three types of distress vocalization elicitors have been delineated: after
72 birth, human infants produce cries when hungry (Gilbert and Robb 1996), when separated
73 from their main caregiver (Christensson et al. 1995), and when in pain (Fuller 1991).

74 Research consisting of testing whether caregivers and experienced professionals can
75 differentiate between cries associated with a diversity of demands and respond accordingly
76 has produced inconclusive results (Wasz-Höckert et al. 1964; Müller et al. 1974). Overall, the
77 current view is that, in humans and other mammals, distress calls represent an acoustically
78 graded system, with no clear acoustic boundaries (Zeskind et al. 1985; Porter et al. 1986;
79 Lingle et al. 2012). However, the fact that a signal is graded does not automatically disqualify
80 it from categorical perception (May et al. 1989; Fischer 1998; Green et al. 2020).

81

82 In this study, we used automated acoustical feature extraction and machine learning to
83 evaluate whether infant chimpanzee distress calling could be associated with specific
84 contextual features (notably: the problem infants experience, the distance to their mother, and
85 the maternal response).

86

87 In chimpanzees, distress calls (typically: ‘whimpers’) are mainly produced by infants (Plooij
88 et al. 1984; Goodall 1986; Bard 2000). Distress calls are different in form and function from
89 alarm calls (typically alarm hoos and barks; see Crockford et al. 2012, 2017, 2018; Schel et al.
90 2013). This difference is already potent in infants and juveniles (Dezecache et al. 2019). One
91 useful functional distinction (at least in chimpanzees) is that distress calls attract social
92 partners to one’s particular situation, whereas alarm calls are used to warn social partners
93 about a common danger in the environment (Dezecache et al., 2019).

94

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

107

108 Recent developments in machine learning have proved critical in the study of animal vocal
109 communication (Mielke and Zuberbühler 2013; Fedurek et al. 2016; Turesson et al. 2016;
110 Versteegh et al. 2016) and human crying (Barajas-Montiel and Reyes-Garcia 2006;

111 Saraswathy et al. 2012; Chang and Li 2016), but classification based on machine learning
112 algorithms has not yet been applied to distress calls of chimpanzees. They can yet help reveal
113 fine differences in acoustical information that discriminates between a number of
114 characteristics of calling episodes, such as the characteristics of the context of calling.

115

116 In order to evaluate the extent to which the distress calls of infant chimpanzees contain
117 information about the situation in which calls occur, the distance between mother and infant,
118 and can be associated with particular maternal responses, we applied automated acoustical
119 information extraction and machine-learning classification methods. We indeed analysed
120 distress call episodes from infants of a cohort of wild chimpanzees (*Pan troglodytes*
121 *schweinfurthii*) (N=8) in the Sonso community of Budongo Forest, Uganda. We extracted and
122 analysed acoustical information from a total of 178 distress call episodes. We subjected
123 acoustic exemplars from a recording sequence to an automated feature extraction algorithm
124 before training a supervised learning algorithm for subsequent categorisation. The procedure
125 consisted of training a model to segregate exemplars given in a particular context. This model
126 was then used to categorise new exemplars (Mohri et al. 2018), which enabled us to evaluate
127 whether the acoustics of graded distress call series encode information about the context of
128 emission. Following prior reports in other primates, we also evaluated whether the distress
129 calls encoded information about the distance between the infant and the mother (Bayart et al.
130 1990; Wiener et al. 1990). Finally, we looked at whether maternal responses could, in
131 principle, be predicted from the acoustic exemplars alone (although other cues are also
132 inevitably used by the mother).

133

134 We predicted that, if distress calls contain information about either the situation and problem
135 triggering infant calling, the distance between mother and infant and can be associated with

136 maternal response, we would observe significant above-baseline level classifications of
137 exemplars unknown to a model previously trained to discriminate exemplars across the
138 various contextual categories.

139

140 Such a discovery could encourage further research to look for a common acoustical code
141 potentially used by the chimpanzee infants and mothers, a code that could contribute to
142 coordinate their activities, notably when parental protective behaviour is most needed.

143

144

145 **Methods**

146 *Ethical note*

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

149

150 *Subjects*

151 Data were collected from infants (N=8) of the well-habituated wild chimpanzees (N≈70) of
152 the Sonso community (Reynolds 2005) of Budongo Forest, Uganda, during February-June
153 2014, December 2014, March-June 2015 and April-May 2016 (see Table 1 for details). For
154 further information about the study site, see Eggeling (1947) and Reynolds (2005). We
155 collected data from those specific infants because they could easily be followed in the forest,
156 as they were born from well-habituated mothers, unwary of human observers.

157

158 *Data collection*

159 Distress calling episodes were continuously videotaped in all-occurrence sampling (Altmann
160 1974). Despite their acoustic variability, the short, tonal, and sometimes sequential production

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

169

170 Infants' perceived needs were classified according to so-called 'situations' ('Separation',
171 'Threat' and 'No threat') and more specific 'problems' using a typology established in the
172 field (see ESM Table 1 for details).

173

174 In some of the 'Separation' situations, physical distance was initiated by the mother and the
175 infant was calling in response to the movement of the mother (problem: 'Active separation';
176 ESM Table 1). In other episodes of the 'Separation' situations, the infant was already away
177 from the mother and started calling in the absence of specific travelling movements from the
178 mother (problem: 'Passive separation'; ESM Table 1).

179

180 Similarly, 'Threat' situations involved different problems. In some cases of 'Threat', infants
181 could be engaged in an activity or action that appeared to trigger physical pain such as due to
182 the mother moving abruptly, or during rough social play (problem: 'Pain'; ESM Table 1). In
183 others, infants found themselves in a threatening environment, such as when there was
184 aggression in the vicinity (problem: 'Danger'; ESM Table 1).

185

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

191

192 Physical distance between mother and infant was determined at the onset of the distress call
193 episodes. We coded ‘supported’ if most of the weight of the infant was then supported by the
194 mother; ‘contact’ if there was physical contact with mother but without full support (e.g.,
195 standing on the ground); ‘arms-reach’ if the infant was within the arm’s reach of the mother,
196 and ‘beyond’ if the infant was beyond the arm’s reach of the mother. Classification of the
197 distances in this manner proved to be more meaningful and probably reliable than the
198 estimation of distance using a metric system (such as measurement in meters or inches). We
199 also determined the nature of the mothers’ reaction starting from the onset of the calling
200 episode until up to 10 seconds after the offset of the calling episode. We coded for whether
201 the mother gazed towards the infant (based on facial orientation), approached the infant,
202 collected the infant or vocalized. Note that these are not mutually exclusive categories in that
203 mothers may, for instance, approach and call simultaneously.

204

205 While the coding of ‘situations’ and ‘problems’ largely relied on live commentaries by the
206 cameraman, the coding of distance and maternal response was mostly based on video. The
207 coding of distance and maternal response was intra-reliably validated on 19.5% of the entire
208 dataset (κ configuration of responses = 0.93; κ distance = 1; κ gaze = 0.90; κ approach = 0.87;
209 κ collection = 0.96; κ vocalization = 0.92).

210

211 *Quantification and statistical analysis*

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

229

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

232

233 *Feature extraction*

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

247

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

254

255 *Feature selection*

256 Prior to classification, we conducted a feature selection procedure by reducing the number of
257 features to a set of reliable features to explain the maximum variance in a given data set. We

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

268

269 *Classification*

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

281

282 We used a radial basis function (RBF) kernel and 5-fold cross-validated the parameters C and
283 Gamma with separate smaller data sets (for details see Fedurek et al. 2016). We used the top
284 40 feature dimensions for classification (see ‘feature selection’ above), omitting all others.
285 We obtained performance scores from the models that were trained and tested on the same
286 labels (basic classification procedure) and from cross-comparisons of conditions, such as
287 training on situation ‘Threat’ and testing on one type of maternal response. Previous work has
288 shown that this method provides useful insights into the nature of information coding
289 (Caldara and Abdi 2006; Fedurek et al. 2016; Dahl et al. 2018).

290

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

301

302 *Feature evaluation*

303 In order to evaluate the features with highest contribution to the classification of certain
304 attributes, we determined the extent to which comparisons shared similar feature dimensions
305 for situations (‘No threat’ vs. ‘Threat’; ‘No threat’ vs. ‘Separation’; ‘Threat’ vs. ‘Separation’).
306 We additionally determined the top 10 feature dimensions of each comparison, as outlined

307 above, and correlated the feature numbers of those feature dimensions in a pairwise fashion
308 using Spearman's rank correlation tests. Feature numbers refer to the topological organization
309 of the mel-frequency space, such that close feature numbers are indicative of similar
310 underlying structures accounting for both comparisons. To determine the feature dimensions
311 that are critical for the classification of exemplars, we assessed whether feature dimensions
312 have been repeatedly used by the classifier overall in the classification. We therefore
313 considered the 15 types of comparisons regarding infants' problems (e.g., 'Conflict' vs 'No
314 reason'). As for the distance between the infant and the mother, we considered the six types of
315 comparisons (e.g., 'Supported' vs. 'Contact'). Regarding maternal response, we considered
316 four yes-no comparisons, namely for 'Gaze', 'Approach', 'Collection' and 'Vocalization'. We
317 then examined the empirical distribution of the 40 feature dimensions used in the feature
318 selection algorithms to determine the top ten contributors. Note that the choice of 10 features
319 was arbitrary. As a baseline, a random distribution of "best features" for each comparison was
320 determined by randomly selecting 10 out of 104 features. The frequency distribution across
321 all comparisons was determined and 95% confidence intervals were calculated by running the
322 procedure 1,000 times. We then reconstructed the underlying frequency bands of significant
323 feature dimensions, resulting in feature maps. In a further step, we calculated differences in
324 feature importance by pairwise contrasting feature maps. We contrasted two feature maps by
325 subtracting corresponding frequency values, reflecting the occurrence of one particular mean
326 cepstrum or co-variance of two cepstra, if at least one of the corresponding two values from
327 the two feature maps was significant.

328

329 **Results**

330 *Classification of exemplars across situations*

331 Infants produced distress calls when separated from their mother (situation: ‘Separation’),
332 when seemingly exposed to threat such as aggression or the experience of pain (situation:
333 ‘Threat’) or for contexts we deemed non-threatening (situation: ‘No threat’). Classification
334 accuracy for all 3 situations was high and significantly higher than baseline (see Figure 1A
335 and ESM Table 2).

336

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

341

342 *Classification of exemplars across problems*

343 The classification accuracy for all problems (but ‘Active separation’ and ‘No reason’) was
344 significantly higher than baseline (see Figure 1B and ESM Table 3). As shown in ESM Figure
345 4A, age did not appear to be driving the observed pattern, as the majority of exemplars were
346 found around the same ages.

347

348 Feature analysis (meant to evaluate the features with highest contributions to the
349 classification) revealed similar rankings for contrasts of ‘Threat’ vs. ‘No threat’ and ‘Threat’
350 vs. ‘Separation’. The correlation of feature numbers at a given rank in both contrasts yielded a
351 positive correlation ($r_s = .93$, $n = 10$, $p < .001$, Figure 1C). This was not the case for pairs of
352 contrasts involving either ‘Separation’ ($r_s = .47$, $n = 10$, $p = .18$) or ‘No threat’ ($r_s = .61$, $n =$

353 10, $p = .07$; Figure 1C). This indicates that the classification between pairs of situations
354 involving ‘Threat’ relied on similar feature dimensions.

355

356 When examining feature dimensions most strongly accounting for the classification of
357 exemplars according to the specific problems the infants were exposed to, we found key
358 contributions from the first, third and fourth mean cepstra, accompanied by covariances of
359 cepstra corresponding with frequency bands around 342.59 to 1074.07 Hz and 50 to 781.48
360 Hz.

361

362 *Classification of exemplars across distances between the infant and the mother*

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

366

367 We contrasted distance classes by assigning categorical units of distance (1 = most proximal
368 [‘Supported’] and 4 = most distal [‘Beyond’]) and by computing the relative distance between
369 classes. Classification accuracy significantly increased as relative distance increased ($F(2,59)$
370 $= 35.12$, $MS = 16.21$, $p < .001$; Figure 2B). This suggests that variation in classification
371 performance reflects coding of distance in the exemplars. The acoustics of calls produced
372 when the relative distance between the mother and the infant was more similar at episode
373 onset were harder to discriminate than those produced when the mother and the infant were
374 relatively more distant to each other.

375

376 Examination of feature dimensions that most contribute to the classification of distances
377 revealed that features that are of significance for two or more comparisons were significantly

378 different from the baseline. Indeed, the empirical number count 2 significantly surpassed a
379 random distribution of number counts, relative to the total number of comparison (here 6)
380 (Figure 2C). The third and fifth mean cepstra, accompanied by covariances of cepstra
381 corresponding with frequency bands in the lower to mid-range of the Hz-spectrum,
382 contributed most to classification (Figure 2D).

383

384 *Classification of exemplars and maternal response*

385 Our model discriminated the probability of the mother to gaze towards, approach, collect the
386 infant and vocalize contingently to the infant's distress calls accurately and above baseline
387 values (see Figure 3A; ESM Table 5). The leave-one-out procedure revealed no significant
388 interaction between individual runs ($N = 8$) and maternal response ($F(3,192) = 0.67$, $MS =$
389 1.87 , $p = 0.570$).

390

391 To evaluate the features most contributing to determining whether the mother did or did not
392 gaze, approach, collected the infant or vocalized contingently to the infant's calling, we
393 plotted the mean expression of each feature dimension for both the presence ('Yes') and the
394 absence ('No') of a given maternal response. Selected features were often located at the outer
395 edge of the distribution (ESM Figure 2B-E), suggesting that they account for one class
396 significantly stronger than for the other class.

397

398 The contrasts that yielded high performance scores showed more distinct feature separation
399 (e.g., 'Gaze' plot - ESM Figure 2B) than the contrasts that yielded lower classification (e.g.,
400 'Collection' plot - ESM Figure 2D). We also examined which feature dimensions accounted
401 for the classification of exemplars the most according to maternal response. This analysis
402 revealed that features that are of significance for two or more comparisons were significantly

403 different from the baseline. The empirical number count 2 significantly surpassed a random
404 distribution of number counts, relative to the total number of comparisons (here 4; Figure 3B).
405 Notably, the first, third, fourth and the 12th mean cepstra, accompanied by accompanied by
406 covariances of cepstra corresponding with frequency bands around 342.59 to 1074.07 Hz and
407 50 to 781.48 Hz, contributed significantly to the classification (Figure 3C). The feature
408 distribution of maternal response classification resembles the feature distribution of problems
409 faced by the infants (Figure 1E).

410

411 *Mother-infant interactions*

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

416

417 Our model's discrimination accuracy was significantly higher than baseline levels for
418 contrasts of the 'Threat' and 'No threat' situations across all maternal response types ($M_{\text{emp}} =$
419 69.38 , $SD_{\text{emp}} = 9.39$; $M_{\text{base}} = 54.65$, $SD_{\text{base}} = 2.87$; $t(78) = 9.48$, $p < 0.001$), as well as of the
420 'Threat' and 'Separation' situations ($M_{\text{emp}} = 72.31$, $SD_{\text{emp}} = 8.87$; $M_{\text{base}} = 55.33$, $SD_{\text{base}} =$
421 2.67 ; $t(78) = 11.57$, $p < 0.001$; Figure 4A). However, contrasts of the other two other
422 situations ('No threat' and 'Separation') were not successfully discriminated based on
423 maternal response ($M_{\text{emp}} = 55.33$, $SD_{\text{emp}} = 4.33$; $M_{\text{base}} = 54.90$, $SD_{\text{base}} = 2.87$; $t(78) = 0.52$, $p =$
424 0.300).

425

426 Testing the model with problems, we found all comparisons to be significant ($p < .05$; ESM
427 Table 6) except for the contrasts of the 'No reason' vs. 'Active separation' problems, and of

428 the ‘Conflict’ vs. ‘No reason’ problems (Figure 4B). Gazing was a relatively reliable action of
429 the mother, particularly during problems associated with Situation ‘Threat’. On the other
430 hand, ‘Collection’ and ‘Vocalization’ were found to be more modest in discriminating
431 between problems.

432

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

448

449 **Discussion**

450 Distress calls have been described as acoustically continuous or graded in various mammals,
451 but studies (notably in humans) have raised the possibility that these calls may convey
452 discrete information (Müller et al. 1974; Wiesenfeld et al. 1981; Brennan and Kirkland 1982;

453 Fuller 1991; Soltis 2004). Here, we used a supervised machine learning approach to
454 investigate distress calls of wild infant chimpanzees and evaluate whether they carry
455 information about the nature of the external events experienced by the caller and whether they
456 can be associated with particular maternal responses, which the calls may elicit.

457

458 We extracted exemplars from distress call episodes and used machine learning classification
459 techniques. We found that a model trained on discriminating exemplars between a threatening
460 situation ('Threat') from others was better at predicting the mother's response, suggesting that
461 distress calls during threatening situations rely on specific acoustic features. This is further
462 exemplified by the fact that discriminations between exemplars found in situations 'Threat'
463 vs. 'Separation' on the one hand, and 'Threat vs. No threat' on the other hand, relied on
464 common features. Our results suggest that the acoustic information present in distress call
465 episodes contain information about the type of external events triggering calling as well as the
466 nature of the problem. In fact, distress calls encode certain information, at the situation level
467 (threat versus no threat), at the problem level, and are associated with maternal responses.
468 These results do not appear to be driven by the inclusion of particular individuals in the
469 dataset (as indicated by insignificant leave-one-out procedures) or by the prominence of
470 particular age classes for specific problems.

471

472 Interestingly, the discrimination of problems that belonged to the same situation (e.g.,
473 problems 'Pain' and 'Danger', both of which occurred in the situation 'Threat') lead to less
474 accurate classification, despite being significantly different from the baseline. The classifier's
475 discrimination of the problems was most accurate when contrasting the problems that
476 occurred in the 'Threat' situation to problems that occurred in the 'No threat' situation.
477 Similarly, contrasts, in which one of the problems occurred in the 'Threat' situation and the

478 other in the ‘Separation’ situation yielded high classification scores. Classification of
479 contrasts involving problems in the ‘Separation’ vs. ‘No threat’ situations appeared to be
480 lower than that comparing problems in the ‘Threat’ vs. other situations. This suggests that
481 threatening contexts are associated with particularly salient acoustical features.

482

483 We also found that the distance between infants and their mothers at the onset of distress call
484 episodes is associated with the particular structure of the exemplars, with exemplars being
485 acoustically most distinctive as the distance between the infants and mothers increases.

486 Another relevant finding was that whether the mothers would gaze at, approach, collect or
487 vocalize at their infant during the call episode could be predicted by the acoustic
488 characteristics of the exemplars. Further, we found that feature maps supporting the
489 classifications of situations and maternal responses were relatively similar.

490

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

500

501 However, it is not clear that mothers solely rely on acoustical cues. A number of other cues
502 (infants’ facial features, movements and gestures; reaction of other conspecifics; knowledge

503 of past problems experienced by the infant) must also be playing a role in shaping mothers'
504 reactions. Our findings only suggest that distress calls encode sufficient information to shape
505 maternal reaction decisions, not that mothers only take acoustic cues into account. Moreover,
506 we cannot firmly establish that infant distress calls are the causes of the maternal responses
507 examined here. It is possible that the mothers were to display the kind of protective response
508 they showed regardless of whether the infant called or not (because they were to do it
509 anyway, or simply because they perceived the problems infants were then facing through
510 other means). The current study does not allow us to disentangle whether, when the mothers
511 responded to their infants' calling, they did it in response to the calls alone or they relied on
512 numerous other cues. A potential way of examining this is by recording distress calls in
513 situations where infants are clearly out of sight of the mothers, a scenario unlikely to happen
514 with young chimpanzee infants, given their relative dependency to their mothers (Plooij et al.
515 1984).

516

517 One strength of the current study is its use of robust machine learning approaches to examine
518 highly complex acoustic phenomena, which would otherwise be intractable. Besides the
519 technical achievement the development of machine learning has undeniably brought in a
520 number of fields of research (Riecken 2000; Sebe et al. 2005; Olsson 2009; Deng and Li
521 2013; Kelleher et al. 2015; Libbrecht and Noble 2015; Nithya and Ilango 2017), its value in
522 the study of animal cognition is becoming clearer (Gerencsér et al. 2013; Wiltschko et al.
523 2015; Dahl et al. 2018). In a recent study, Fedurek and colleagues (Fedurek et al. 2016) were
524 able to show that different phases of the complex pant-hoot calls of male chimpanzees carry
525 information about particular features of the caller, notably its age and identity, revealing
526 unprecedented details about how information can be acoustically represented.

527

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

535

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

545

546 **References**

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

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

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

554 Bard KA (2000) Crying in infant primates: Insights into the development of crying in
555 chimpanzees. In: *Crying as a Sign, a Symptom, and a Signal: Clinical, Emotional and*

- 556 Developmental Aspects of Infant and Toddler Crying. Mac Keith Press, London, p
557 157
- 558 Bayart F, Hayashi KT, Faull KF, et al (1990) Influence of maternal proximity on behavioral
559 and physiological responses to separation in infant rhesus monkeys (*Macaca mulatta*).
560 Behavioral neuroscience 104:98
- 561 Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and
562 powerful approach to multiple testing. Journal of the Royal statistical society: series B
563 (Methodological) 57:289–300
- 564 Boll S (1979) Suppression of acoustic noise in speech using spectral subtraction. IEEE
565 Transactions on acoustics, speech, and signal processing 27:113–120
- 566 Brennan M, Kirkland J (1982) Classification of infant cries using descriptive scales. Infant
567 Behavior and Development 5:341–346
- 568 Briefer EF (2012) Vocal expression of emotions in mammals: mechanisms of production and
569 evidence. Journal of Zoology 1–20. [https://doi.org/10.1111/j.1469-
570 7998.2012.00920.x@10.1111/\(ISSN\)1469-7998.SoundProduction](https://doi.org/10.1111/j.1469-7998.2012.00920.x@10.1111/(ISSN)1469-7998.SoundProduction)
- 571 Caldara R, Abdi H (2006) Simulating the ‘other-race’ effect with autoassociative neural
572 networks: further evidence in favor of the face-space model. Perception 35:659–670
- 573 Chang C-C, Lin C-J (2011) LIBSVM: A library for support vector machines. ACM
574 transactions on intelligent systems and technology (TIST) 2:1–27
- 575 Chang C-Y, Li J-J (2016) Application of deep learning for recognizing infant cries. In: 2016
576 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW). IEEE,
577 pp 1–2
- 578 Choqueuse V (2020) Vuvuzela sound denoising algorithm. Version 1.2.0 URL
579 [https://fr.mathworks.com/matlabcentral/fileexchange/27912-vuvuzela-sound-
580 denoising-algorithm?focused=b421486c-db15-d58a-6f4d-
581 5026e5f66319&tab=function](https://fr.mathworks.com/matlabcentral/fileexchange/27912-vuvuzela-sound-denoising-algorithm?focused=b421486c-db15-d58a-6f4d-5026e5f66319&tab=function)
- 582 Christensson K, Cabrera T, Christensson E, et al (1995) Separation distress call in the human
583 neonate in the absence of maternal body contact. Acta paediatrica 84:468–473
- 584 Crockford C, Gruber T, Zuberbühler K (2018) Chimpanzee quiet hoo variants differ
585 according to context. Royal Society open science 5:172066
- 586 Crockford C, Wittig RM, Mundry R, Zuberbühler K (2012) Wild chimpanzees inform
587 ignorant group members of danger. Current Biology 22:142–146.
588 <https://doi.org/10.1016/j.cub.2011.11.053>
- 589 Crockford C, Wittig RM, Zuberbühler K (2017) Vocalizing in chimpanzees is influenced by
590 social-cognitive processes. Sci Adv 3:e1701742.
591 <https://doi.org/10.1126/sciadv.1701742>
- 592 Dahl CD, Wyss C, Zuberbühler K, Bachmann I (2018) Social information in equine
593 movement gestalts. Animal cognition 21:583–594

- 594 Deng L, Li X (2013) Machine learning paradigms for speech recognition: An overview. *IEEE*
595 *Transactions on Audio, Speech, and Language Processing* 21:1060–1089
- 596 Dezecache G, Crockford C, Zuberbühler K (2019) The development of communication in
597 alarm contexts in wild chimpanzees. *Behav Ecol Sociobiol* 73:104.
598 <https://doi.org/10.1007/s00265-019-2716-6>
- 599 Eggeling WJ (1947) Observations on the ecology of the Budongo rain forest, Uganda. *The*
600 *Journal of Ecology* 20–87
- 601 Ephraim Y, Malah D (1984) Speech enhancement using a minimum-mean square error short-
602 time spectral amplitude estimator. *IEEE Transactions on acoustics, speech, and signal*
603 *processing* 32:1109–1121
- 604 Fedurek P, Zuberbühler K, Dahl CD (2016) Sequential information in a great ape utterance.
605 *Scientific reports* 6:38226
- 606 Fischer J (1998) Barbary macaques categorize shrill barks into two call types. *Animal*
607 *behaviour* 55:799–807
- 608 Fröhlich M, Hobaiter C (2018) The development of gestural communication in great apes.
609 *Behavioral Ecology and Sociobiology* 72:194
- 610 Fröhlich M, Wittig RM, Pika S (2016) Should I stay or should I go? Initiation of joint travel
611 in mother–infant dyads of two chimpanzee communities in the wild. *Animal*
612 *Cognition* 19:483–500
- 613 Fröhlich M, Wittig RM, Pika S (2019) The ontogeny of intentional communication in
614 chimpanzees in the wild. *Developmental science* 22:e12716
- 615 Fuller BF (1991) Acoustic discrimination of three types of infant cries. *Nursing Research*
616 40:156–160
- 617 Gerencsér L, Vásárhelyi G, Nagy M, et al (2013) Identification of behaviour in freely moving
618 dogs (*Canis familiaris*) using inertial sensors. *PloS one* 8:e77814
- 619 Gilbert HR, Robb MP (1996) Vocal fundamental frequency characteristics of infant hunger
620 cries: birth to 12 months. *International journal of pediatric otorhinolaryngology*
621 34:237–243
- 622 Goodall J (1986) *The chimpanzees of Gombe: Patterns of behavior*. Harvard University Press
- 623 Green PA, Brandley NC, Nowicki S (2020) Categorical perception in animal communication
624 and decision-making. *Behav Ecol*. <https://doi.org/10.1093/beheco/araa004>
- 625 Hobaiter C, Byrne RW (2011) The gestural repertoire of the wild chimpanzee. *Animal*
626 *cognition* 14:745–767
- 627 Hobaiter C, Byrne RW (2014) The meanings of chimpanzee gestures. *Current Biology*
628 24:1596–1600

- 629 Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian journal*
630 *of statistics* 65–70
- 631 Illingworth RS (1955) Crying in Infants and Children. *Br Med J* 1:75–78
- 632 Kelleher JD, Mac Namee B, D’arcy A (2015) *Fundamentals of machine learning for*
633 *predictive data analytics: algorithms, worked examples, and case studies*. MIT press
- 634 Levréro F, Mathevon N (2013) Vocal signature in wild infant chimpanzees. *American Journal*
635 *of Primatology* 75:324–332
- 636 Libbrecht MW, Noble WS (2015) Machine learning applications in genetics and genomics.
637 *Nature Reviews Genetics* 16:321–332
- 638 Lingle S, Riede T (2014) Deer mothers are sensitive to infant distress vocalizations of diverse
639 mammalian species. *The American Naturalist* 184:510–522
- 640 Lingle S, Wyman MT, Kotrba R, et al (2012) What makes a cry a cry? A review of infant
641 distress vocalizations. *Current Zoology* 58:698–726
- 642 Logan B (2000) Mel frequency cepstral coefficients for music modeling. In: *Ismir*. pp 1–11
- 643 Lowe AE, Hobaiter C, Asiimwe C, et al (2020) Intra-community infanticide in wild, eastern
644 chimpanzees: a 24-year review. *Primates* 61:69–82
- 645 Lowe AE, Hobaiter C, Newton-Fisher NE (2019) Countering infanticide: chimpanzee
646 mothers are sensitive to the relative risks posed by males on differing rank trajectories.
647 *American journal of physical anthropology* 168:3–9
- 648 MacLean PD (1985) Brain evolution relating to family, play, and the separation call. *Archives*
649 *of general psychiatry* 42:405–417
- 650 May B, Moody DB, Stebbins WC (1989) Categorical perception of conspecific
651 communication sounds by Japanese macaques, *Macaca fuscata*. *The Journal of the*
652 *Acoustical Society of America* 85:837–847. <https://doi.org/10.1121/1.397555>
- 653 Mielke A, Zuberbühler K (2013) A method for automated individual, species and call type
654 recognition in free-ranging animals. *Animal Behaviour* 86:475–482
- 655 Mohri M, Rostamizadeh A, Talwalkar A (2018) *Foundations of machine learning*. MIT press
- 656 Müller E, Hollien H, Murry T (1974) Perceptual responses to infant crying: Identification of
657 cry types. *Journal of Child Language* 1:89–95
- 658 Newman JD (2007) Neural circuits underlying crying and cry responding in mammals.
659 *Behavioural brain research* 182:155–165
- 660 Nithya B, Ilango V (2017) Predictive analytics in health care using machine learning tools
661 and techniques. In: *2017 International Conference on Intelligent Computing and*
662 *Control Systems (ICICCS)*. IEEE, pp 492–499

- 663 Olsson F (2009) A literature survey of active machine learning in the context of natural
664 language processing
- 665 Plooij FX, Bygott D, Sackett GP (1984) The behavioral development of free-living
666 chimpanzee babies and infants. Ablex Norwood, NJ
- 667 Porter FL, Miller RH, Marshall RE (1986) Neonatal pain cries: effect of circumcision on
668 acoustic features and perceived urgency. *Child development* 790–802
- 669 Reynolds V (2005) The chimpanzees of the Budongo forest: Ecology, behaviour and
670 conservation. OUP Oxford
- 671 Riecken D (2000) Introduction: personalized views of personalization. *Communications of*
672 *the ACM* 43:26–28
- 673 Saraswathy J, Hariharan M, Yaacob S, Khairunizam W (2012) Automatic classification of
674 infant cry: A review. In: 2012 International Conference on Biomedical Engineering
675 (ICoBE). pp 543–548
- 676 Schel AM, Townsend SW, Machanda Z, et al (2013) Chimpanzee alarm call production meets
677 key criteria for intentionality. *PloS one* 8:e76674.
678 <https://doi.org/10.1371/journal.pone.0076674>
- 679 Sebe N, Cohen I, Garg A, Huang TS (2005) Machine learning in computer vision. Springer
680 Science & Business Media
- 681 Soltis J (2004) The signal functions of early infant crying. *Behavioral and brain sciences*
682 27:443–458
- 683 Tajiri Y, Yabuwaki R, Kitamura T, Abe S (2010) Feature extraction using support vector
684 machines. In: *International Conference on Neural Information Processing*. Springer,
685 pp 108–115
- 686 Townsend SW, Slocombe KE, Thompson ME, Zuberbühler K (2007) Female-led infanticide
687 in wild chimpanzees. *Current Biology* 17:R355–R356
- 688 Turesson HK, Ribeiro S, Pereira DR, et al (2016) Machine learning algorithms for automatic
689 classification of marmoset vocalizations. *PloS one* 11:e0163041
- 690 Versteegh M, Kuhn J, Synnaeve G, et al (2016) Classification and automatic transcription of
691 primate calls. *The Journal of the Acoustical Society of America* 140:EL26–EL30
- 692 Wasz-Höckert O, Partanen TJ, Vuorenkoski V, et al (1964) The identification of some
693 specific meanings in infant vocalization. *Experientia* 20:154–154
- 694 Watts DP, Mitani JC (2000) Infanticide and cannibalism by male chimpanzees at Ngogo,
695 Kibale National Park, Uganda. *Primates* 41:357–365
- 696 Weary DM, Lawson GL, Thompson BK (1996) Sows show stronger responses to isolation
697 calls of piglets associated with greater levels of piglet need. *Animal Behaviour*
698 52:1247–1253. <https://doi.org/10.1006/anbe.1996.0272>

- 699 Wiener SG, Bayart F, Faull KF, Levine S (1990) Behavioral and physiological responses to
700 maternal separation in squirrel monkeys (*Saimiri sciureus*). Behavioral neuroscience
701 104:108
- 702 Wiesenfeld AR, Malatesta CZ, Deloach LL (1981) Differential parental response to familiar
703 and unfamiliar infant distress signals. Infant Behavior and Development 4:281–295
- 704 Wiltschko AB, Johnson MJ, Iurilli G, et al (2015) Mapping sub-second structure in mouse
705 behavior. Neuron 88:1121–1135
- 706 Zeskind PS, Sale J, Maio ML, et al (1985) Adult perceptions of pain and hunger cries: A
707 synchrony of arousal. Child development 549–554
- 708

709 **Tables**

710 **Table 1**

711 List of infants, estimated birthdate, minimum and maximum age in months over the course of
 712 the study, sex, number of contributed distress call episodes used in acoustic extraction, and
 713 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

714

715

716 **Figures**

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

719 (A, B) Accuracy values of classification runs are shown for all contrasts between situations
720 (panel A) and problems (panel B). Wide blue horizontal bars indicate the means and blue
721 horizontal bars the 95% confidence-intervals. Wide red lines indicate the means of the
722 baseline and narrow red lines the 95% confidence-intervals. (C) The feature sets accounting
723 for individual contrasts the most are compared and correlated. Dots show individual samples;
724 relationships between them are indicated by regression lines. (D) Feature evaluation
725 procedure, showing the occurrence count (x-axis), reflecting the number of times a particular
726 feature dimension was among the top 10 feature dimensions across all comparisons, and the
727 relative frequency of n-counts (y-axis). The empirical distribution is shown in red, a random
728 distribution (solid line) in black, and 95% confidence-interval in dotted lines. N-counts of 4 or
729 more are significantly over-represented. (E) X- and Y-axes represent frequency bands.
730 Significant feature dimensions are colour-coded and structurally aligned in a frequency-
731 transformed representation. Light blue dots indicate significant mean cepstra, orange dots
732 indicate positive co-variances of cepstra, and mint dots indicate negative co-variances of
733 cepstra. Grey dots indicate non-significant feature dimensions. The size of dots indicates their
734 relative importance: the larger the dot the more frequently a feature dimension has been used
735 across all comparisons.

736

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

739 (A) Accuracy values of classification runs are shown, contrasting various categorical
740 distances between the infant and the mother. The colour-code is the same as in Figure 1

741 panels A & B. (B) Accuracy scores are plotted according to the relative distances compared
742 (max 3: Supported [Supported] - Beyond arm's reach [Beyond]). Accuracy values are shown
743 as z-scores. Red line represents the linear regression fit. (C) Feature evaluation procedure,
744 showing the occurrence count (x-axis), reflecting the number of times a particular feature
745 dimension was among the top 10 feature dimensions across all comparisons, and the relative
746 frequency of n-counts (y-axis). The colour-code is the same as in Figure 1D. N-counts of 2 or
747 more are significantly over-represented. (D) Significant feature dimensions (see Figure 1E for
748 colour-codes).

749

750 ***Figure 3. Classification performance per type of maternal response and evaluation of***
751 ***contributing features***

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

759

760 ***Figure 4. Mother-infant interaction classification models***

761 (A & B) Accuracy values of classification runs for models trained on situation (A) or infants'
762 problems (B) contrasts (y-axis) and tested on maternal response contrasts (x-axis). The colour
763 represents the level of statistical significance. (C, D, E) Feature map comparison. Feature
764 maps were contrasted by subtracting corresponding frequency values, reflecting the
765 occurrence of one particular mean cepstra or co-variance of two cepstra, if at least one of the

766 corresponding two values from the two feature maps was significant. Mint indicates greater
767 importance for the first term (e.g., 'situation' in the extreme-left panel), orange for the second
768 term of the comparison (e.g., 'contingent maternal behaviour' in the extreme-right panel).















