

1 **Humans recognize affective cues in primate vocalizations:**
2 **Acoustic and phylogenetic perspectives**

3
4
5 Debracque, C.^{1*}, Clay, Z.², Grandjean, D.^{1±}, & Gruber, T.^{1±}
6
7

8 ¹Department of Psychology and Educational Sciences and Swiss Center for Affective Sciences,
9 University of Geneva, Geneva, Switzerland;

10 ²Department of Psychology, Durham University, Durham, United Kingdom;

11 [±]*joint co-senior authors*
12

13 *Correspondence to: Coralie.Debracque@unige.ch; Swiss Center for Affective Sciences,
14 Campus Biotech, CISA - University of Geneva, Chemin des Mines 9, 1202, Geneva,
15 Switzerland; tel: +41223790889

16 **Abstract**

17 Humans are adept in extracting affective information from the vocalisations of not only humans
18 but also other animals. Current research has mainly focused on phylogenetic proximity to
19 explain such cross-species emotion recognition abilities. However, because research protocols
20 are inconsistent across studies, it remains unclear whether human recognition of vocal affective
21 cues of other species is due to cross-taxa similarities between acoustic parameters, the
22 phylogenetic distances between species, or a combination of both. To address this, we first
23 analysed acoustic variation in 96 affective vocalizations, including agonistic and affiliative
24 contexts, of humans and three other primate species – rhesus macaques, chimpanzees and
25 bonobos – the latter two being equally phylogenetically distant from humans. Using
26 Mahalanobis distances, we found that chimpanzee vocalizations were acoustically closer to
27 those of humans than to those of bonobos, confirming a potential derived vocal evolution in
28 the bonobo lineage. Second, we investigated whether 68 human participants recognized the
29 affective basis of vocalisations through tasks by asking them to categorize ('A vs B') or
30 discriminate ('A vs non-A') vocalisations based on their affective content. Results showed that
31 participants could reliably categorize and discriminate most of the affective vocal cues
32 expressed by other primates, except threat calls by bonobos and macaques. Overall, participants
33 showed greatest accuracy in detecting chimpanzee vocalizations; but not bonobo vocalizations,
34 which provides support for both the phylogenetic proximity and acoustic similarity hypotheses.
35 Our results highlight for the first time the importance of both phylogenetic and acoustic
36 parameter level explanations in cross-species affective perception, drawing a more complex
37 picture to explain our natural understanding of animal signals.

38

39 **Keywords**

40 Categorization, discrimination, affect, vocalization, primate, acoustic, phylogeny

41

42

43

44

45

46

47

48

49

50

51 Introduction

52 Vocal communication of affect is crucial for the emotional and attentional regulation of human
53 social interactions (Grandjean et al., 2005; Sander et al., 2005; Schore & Schore, 2008). For
54 instance, the modulation of prosodic features in human speech such as intonation or amplitude
55 can convey subtle affective information to receivers (Grandjean, Bänziger, & Scherer, 2006;
56 Scherer, 2003). Humans consistently recognize and evaluate the affective cues of others' vocal
57 signals in tasks with varying levels of complexity, with emotion categorization i.e. unbiased
58 choice (A versus B) seemingly more cognitively complex than discrimination i.e. biased choice
59 (A versus non-A) (Dricu et al., 2017; Gruber et al. 2020). In both emotion categorization and
60 discrimination tasks, research shows that listeners can subjectively attribute the speaker's
61 reported affective state (i.e. angry, fearful or happy) as well as any potentially referential
62 content (Brunswick, 1956; Grandjean et al., 2006). By no means uniquely human, these
63 affective identification mechanisms facilitate adaptive behaviour in animals such as to
64 approach or avoid the stimulus (Frijda, 1987, 2016; Gross, 1998; Nesse, 1990). Hence, current
65 mechanisms underlying human and other animal vocalizations seem to result from similar
66 adaptive pressures. For instance, research has shown the critical role of acoustic roughness in
67 both human and great ape fear screams to rapidly appraise danger (Arnal et al., 2015; Kret et
68 al., 2020). Despite the adaptive value and importance of auditory affective processing to our
69 own species, its evolutionary origins remain poorly understood.

70
71 As noted, the adaptive behaviours underpinning communication of affect are often shared
72 amongst animals. Over a century ago, Darwin (1872) hypothesized an evolutionary continuity
73 between human and other animals for the vocal expression of affective signals. Morton (1977,
74 1982) subsequently proposed a model of motivational structural rules to characterize the
75 relationship between the acoustic structure of mammal and bird vocalizations and their
76 presumed affective contents. The systematic modulation of call acoustic structure and the
77 caller's underlying affective state appear to provide reliable cues that allow listeners to evaluate
78 aspects of the eliciting stimulus, such as the level of threat or danger (Anderson & Adolphs,
79 2014; Filippi et al., 2017). Comparative research has confirmed that conspecifics are sensitive
80 to such cues, with playback studies showing that both chimpanzees and rhesus macaques
81 discriminate between agonistic screams produced by victims facing varying degrees of threat
82 (Slocombe, Townsend, & Zuberbühler, 2009; Gouzoules, 1984), while meerkats extrapolate
83 the degree of urgency required from the acoustic structure of conspecific alarm calls (Manser,
84 2001). This evidence suggests an evolutionary continuity in the vocal processing ability of both
85 humans and non-human primates to accurately identify affective cues in conspecific
86 vocalizations (Gruber & Grandjean, 2017).

87
88 Interestingly, this evolutionary continuity is also suggested by a second line of research, which
89 shows that human participants generally perform above chance asked to identify primate
90 signals. Despite a limited number of currently available studies (eight, to our knowledge -
91 Belin, Fecteau, et al., 2008; Ferry et al., 2013; Filippi et al., 2017; Fritz et al., 2018; Kamiloğlu
92 et al., 2020; Kelly et al., 2017; Linnankoski et al., 1994; Scheumann et al., 2014, 2017), existing
93 findings on human perception of arousal and valence in non-human primate calls are

94 promising. Indeed, research has shown that humans can discriminate the valence of
95 chimpanzee vocalizations, including agonistic screams (negative valence) and food-associated
96 calls (positive valence) (Fritz et al., 2018; Kamiloğlu et al., 2020); by comparison however,
97 behavioural discrimination for rhesus macaque calls given in the same contexts is poor (Fritz
98 et al., 2018; Belin et al., 2008). Functional Magnetic Resonance Imaging (fMRI) measures
99 taken by Fritz and collaborators also showed that neural activations were more similar when
100 attending to chimpanzee and human vocalizations than macaque calls. In contrast, Linnankoski
101 and colleagues (1994) found that both human adults and infants could categorize affective
102 macaque vocalizations in a larger range of contexts (angry, fearful, satisfied, scolding and
103 submissive). Methodological differences might explain the differences in previous findings
104 concerning macaque calls: it may be easier for human adults and infants to label affective
105 contents of non-human primate vocalizations in a forced choice paradigm (categorization or
106 discrimination tasks) in which the number of possibilities is limited rather than to rate the
107 valence or arousal using Likert scales. For instance, research with human affective stimuli
108 using forced choice paradigms demonstrated the positive relationship between cognitive
109 complexity and the number of available categories to choose from (Dricu et al., 2017; Gruber
110 et al. 2020). Thus, forced choice paradigms with limited options to choose from may lead to
111 elevated performance with macaque calls (Linnankoski et al., 1994) compared to paradigms
112 with Likert rating scales (Belin et al., 2008; Fritz et al., 2018).

113
114 In addition to the mixed findings concerning human sensitivity to valence in non-human
115 primate vocalisations, evidence that humans can accurately judge vocal arousal in other species
116 is also mixed. Recent findings highlight the ability of humans to reliably identify arousal in
117 barbary macaque vocalizations expressed in negative contexts (Filippi et al., 2017) and arousal
118 ratings of chimpanzee vocalizations seem to be fairly accurate across positive and negative
119 valences (Kamiloğlu et al., 2020). Yet, Kelly and collaborators (2017) also showed that human
120 participants over-estimated the distress content of bonobo infant calls compared to those of
121 human or chimpanzee ones, suggesting a relatively poor capacity of humans to identify arousal
122 in bonobo vocalizations. Overall, humans appear to perform relatively well with chimpanzee
123 calls (Kamiloğlu et al., 2020), but less well with bonobo or macaque calls. However, it remains
124 unclear why this is the case. In addition, it is also relevant to examine why a particular primate
125 species, human especially, may be able to recognize affective vocalizations expressed by
126 another primate species.

127
128 Several factors might explain our abilities to recognize some species' affective vocalizations
129 more reliably than others. Previous studies comparing human responses to closely and distantly
130 related species, have highlighted the importance of phylogenetic proximity in human
131 recognition of affect (e.g. Belin et al. 2008, Fritz et al. 2018), arguing that we are more sensitive
132 to emotional content of vocalisations in closely related species. An important test of this
133 hypothesis is to examine responses to vocalisations of two species that are equally closely
134 related to humans. Only one study has attempted this to date by comparing human responses
135 to chimpanzee and bonobo vocalisations, humans closest living relatives (Gruber & Clay,
136 2016). Focusing on distress calls, Kelly et al (2017) found that humans were less accurate at
137 rating distress intensity in bonobo calls compared to chimpanzee calls, but whether this pattern

138 generalizes beyond distress calls is currently unknown.

139

140 In addition to phylogenetic proximity, another important factor determining human accuracy
141 at detecting the emotional content of other species vocalisations may be similarity in the
142 acoustic parameters of vocalisations between humans and the test species. Previous studies
143 have revealed cross-taxa similarities in the acoustic conveyance of affect (Ross, Owren, &
144 Zimmermann, 2009; Scheumann et al., 2014). In particular, previous research has linked the
145 human ability to recognize affective cues from vocalizations of other species to specific
146 modulations of the fundamental frequency (F0), the mean pitch or the energy of the affective
147 calls expressed by non-human primates (Briefer, 2012, Filippi et al., 2017; Linnankoski et al.,
148 1994; Scheumann et al., 2014). Concurrently, acoustic similarity is also influenced by the call's
149 emotional valence (Belin, Fecteau, et al., 2008). Despite being as equally related to us as
150 chimpanzees, the vocal repertoire of bonobos shows some notable acoustic differences,
151 including elevated pitch (Tuttle, 1993) potentially due to shorter vocal tracts (Grauwunder et
152 al 2018). Hence, it seems reasonable to hypothesize that acoustic differences in bonobo calls
153 may lead to lower performance in a human recognition task.

154

155 Overall, it thus remains unclear whether the human ability to recognize affective vocal cues
156 from other species is mainly due to (1) cross-taxa similarities in acoustic parameters, (2) the
157 phylogenetic distances between species, or (3) both, considering that closely phylogenetically-
158 related species may be likely to share acoustic parameters. To address these outstanding issues,
159 we designed a forced-choice paradigm, where participants had to perform two tasks:
160 categorization (A versus B, cognitively demanding) and discrimination (A versus non-A; less
161 cognitively demanding). In both tasks, participants had to judge the affective nature of
162 vocalisations produced in three affective contexts (threat, distress and affiliation) by humans
163 and three other primate species that vary in phylogenetic distance to humans (equally close to
164 humans: chimpanzee, bonobo; more distant: rhesus macaque). For each of the two tasks we
165 measured whether participants were significantly above chance, and whether accuracy of
166 performance could be predicted by species, affect or their interaction. To disentangle whether
167 human cross-species emotion recognition performance was best explained by phylogenetic
168 distance or acoustic similarity, we first established the acoustic similarity of chimpanzee,
169 bonobo and macaque vocalisations to human vocalisations. We calculated Mahalanobis
170 distances to compare the acoustic distances between vocalizations of various affective contexts
171 from these species. We expected that if phylogenetic distance was the main determinant of
172 performance, recognition of affective cues in human vocalisation should be greater than those
173 of chimpanzees and bonobos, which should be equally better than those of rhesus monkey
174 vocalizations (Humans>Chimpanzees=bonobos>macaques). By contrast, if acoustic similarity
175 was the main determinant of performance, participants should perform best with the calls of
176 species most acoustically similar to those of humans. If we found a significant interaction
177 between species and affect on Mahalanobis distance of calls to the human centroid, then
178 recognition performance would need to be compared to acoustic similarity between species at
179 the level of each affect. Moreover, both phylogenetic proximity and acoustic distance may both
180 play a role in explaining human cross species emotional recognition. We may expect amongst
181 equally related species, more accurate performance with the species most similar acoustic

182 structures to humans (if chimpanzees are shown to be more acoustically similar to humans than
183 bonobos overall, or for certain affects, we might expect better recognition accuracy for
184 chimpanzees than bonobos: Humans > Chimpanzees > Bonobos > Macaques). Finally, because
185 of the previous literature (Dricu et al. 2017; Gruber et al. 2020), we also expected participants
186 to perform more accurately on discrimination rather than categorisation tasks.

187

188 **Materials and methods**

189 *Participants*

190 Sixty-eight healthy adult volunteers from the Geneva area (29 males; mean age 23.54 years,
191 SD = 5.09, age range 20 – 37 years) took part in the experiment. The participants reported
192 normal hearing abilities and normal or corrected-to-normal vision. No participant presented a
193 neurological or psychiatric history, or a hearing impairment. All participants gave informed
194 and written consent for their participation in accordance with the ethical and data security
195 guidelines of the University of Geneva. The study was approved by the Ethics Cantonal
196 Commission for Research of the Canton of Geneva, Switzerland (CCER).

197

198 *Vocal stimuli*

199 For our stimuli, we compiled a set of ninety-six vocalizations balanced across four primate
200 species (human, chimpanzee, bonobo, rhesus macaque) and three affective contexts (threat,
201 distress and affiliation). For human stimuli, non-linguistic vocal stimuli from two male and two
202 female actors denoted as expressing a happy, angry or fearful affect were obtained from the
203 Montreal Affective Voices Audio Collection (Belin, Fillion-Bilodeau, et al., 2008). For
204 chimpanzee, bonobo and rhesus macaque stimuli, vocalizations taken from existing author
205 databases were compiled from corresponding contexts: affiliation - food-associated grunts,
206 threat - aggressor barks in agonistic contexts, and distress calls - victims in social conflicts. For
207 each species, 24 stimuli taken from 6-8 different individuals were selected containing single
208 calls or two call sequences of a single individual. All vocal stimuli were standardized to 750
209 milliseconds using PRAAT (www.praat.org) but were not normalized for energy to preserve
210 the naturalness of the sounds (Ferdenzi et al., 2013).

211 *Experimental procedure*

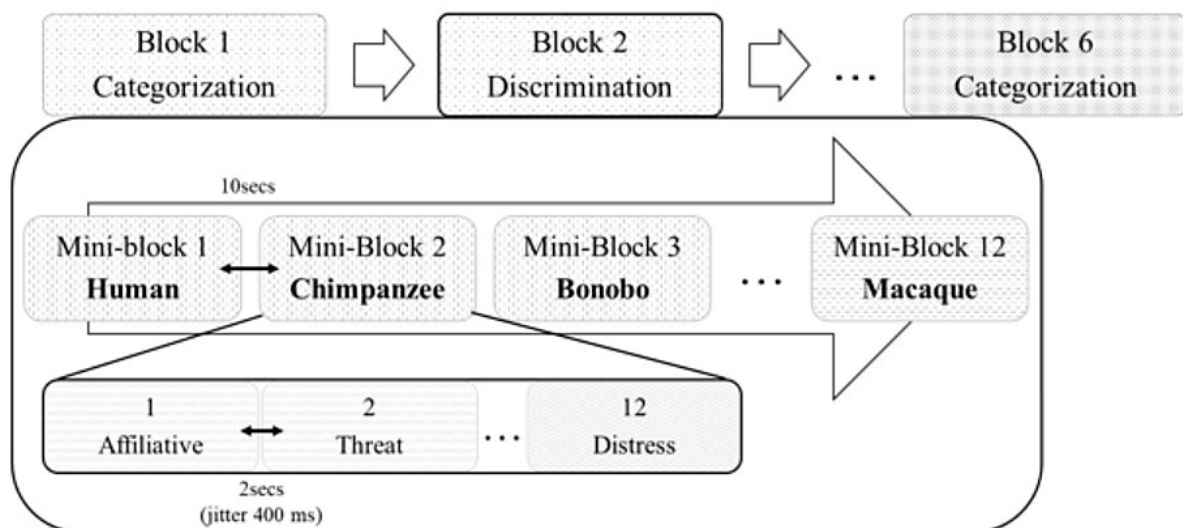
212 Seated in front of a computer, participants listened to the vocalizations played binaurally using
213 Seinnheiser headphones at 70 dB SPL. Each of the 96 stimuli was repeated nine times across
214 six separate counterbalanced blocks leading to 864 trials following a randomization process.
215 The overall experiment followed a within-subjects design with various layers (Figure 1).
216 Testing blocks were task-specific, with participants either performing a categorization task (A
217 versus B) or a discrimination task (A versus non-A). Participants completed three
218 categorization blocks and three discrimination blocks, resulting in six blocks in total. Each
219 block was made of 12 mini-blocks, each separated by a break of 10 seconds. Mini-blocks
220 comprised one unique mini-block per species (human, chimpanzee, bonobo and rhesus

221 macaque), each mini-block repeated 3 times. Within each mini-block were 12 trials, containing
222 four vocalisations from all three contexts (affiliative/happy; threatening/anger; distress/fear)
223 produced by a single species. The blocks, mini-blocks and stimuli were pseudo-randomly
224 assigned for each participant to avoid more than two consecutive blocks, mini-blocks and
225 stimuli from the same category.

226

227 At the beginning of each block, participants were instructed to identify the affective content of
228 the vocalizations using a keyboard. For instance, the instructions for the categorization task
229 could be “Affiliative – press M or Threatening – press Z or Distress – press space bar”.
230 Similarly, the instructions for discrimination could be “Affiliative – press Z or other affect –
231 press M”. The pressed keys were randomly assigned across blocks and participants. The
232 participants pressed the key during 2-second intervals (jittering of 400 ms) between each
233 stimulus. If the participant did not respond during this interval, the next stimulus followed
234 automatically.

235



236

237

238 Figure 1: Structure of the experiment, with each of the six blocks made of 12 mini-blocks,
239 which in turn comprised 12 individual trials.

240

241 *Statistical analysis*

242 **Acoustic analyses**

243 To quantify the impact of acoustic distance in human affect recognition of primate
244 vocalizations, we automatically extracted 88 acoustic parameters from all stimuli vocalizations
245 using the extended Geneva Acoustic parameters set, which is defined as the optimal acoustic
246 indicators related to human voice analysis (GeMAPS; Eyben et al., 2016). This set of acoustical
247 parameters was selected based on i) their potential to index affective physiological changes in
248 voice production, ii) their proven value in former studies as well as their automatic
249 extractability, and iii) their theoretical significance. This set of acoustic parameters includes

250 related frequency parameters (e.g. pitch, jitter, formants), energy parameters (e.g. loudness,
251 shimmer), and spectral parameters (e.g. alpha ratio, Hammarberg index, spectral slopes).

252

253 To assess the acoustic distance between vocalizations of all species, we then ran a Discriminant
254 Analysis (DA) using SPSS 26.0.0.0 based upon the 88 acoustical parameters in order to
255 discriminate our stimuli based on the four different species (human, chimpanzee, bonobo, and
256 rhesus macaque). Excluding the acoustical variables with the highest correlations ($>.90$) to
257 avoid redundancy of acoustic parameters, we retained 16 acoustic parameters related to
258 frequency, energy, and spectral parameters that could discriminate species (see Supplementary
259 material Table S1).

260

261 Using these 16 acoustic features, we subsequently computed Mahalanobis distances of the 96
262 experimental stimuli. A Mahalanobis distance is obtained from a generalized pattern analysis
263 computing the distance of each vocalization from the centroids of the different species
264 vocalizations (Mahalanobis, 1936). This analysis allowed us to obtain an acoustical distance
265 matrix used to test how these acoustical distances were differentially related to the different
266 species. To test this, we performed Generalized Linear Mixed Models (GLMMs) fitted by
267 Restricted Maximum Likelihood (REML) on R.studio (Team, 2020) using the package Lme4
268 (Bates et al., 2015) to test whether the following three fixed factors could predict the
269 Mahalanobis distances: Species (the species which produced the vocalization), Distance-
270 Species (the species centroid used to compute the distance for the same species or for the other
271 species, e.g. the human centroid used to quantify the distance of chimpanzee vocalization from
272 humans), and Affect (affiliative, threat, and distress). We also examined the interaction
273 between these three factors. The identity of the vocalizer was included as a random factor.

274

275 To test the effects of phylogenetic distance, we performed contrasts of interest on the factor of
276 Species (i.e. human < chimpanzee=bonobo < macaque) taking into account the other fixed and
277 random factors. In order to identify the acoustic similarity between human vocalisations and
278 those of chimpanzees, bonobos and macaques, we performed relevant pairwise comparisons
279 on Mahalanobis distances from the centroid of Human vocalizations: for each affect, we
280 compared: Human vs Chimpanzee, Human vs Bonobo; Human vs Macaque; Chimpanzee vs
281 Bonobo; Chimpanzee vs Macaque and Bonobo vs Macaque. Hence, each subset of data (e.g.
282 threat chimpanzee) appeared a maximum total of 3 times in the pairwise comparisons, leading
283 us to compare our p-values to Bonferroni corrected alpha level of $P_{\text{corrected}} = .05/3 = .017$.

284

285 **Vocal recognition performance**

286 First, we investigated if participants' recognition accuracy in the categorisation and
287 discrimination tasks was significantly above chance for each affect per species (i.e. three affects
288 x 4 species = 12 separate tests). Per participant, we calculated the proportion of correct answers
289 for each affect-species set of calls ($N = 8$ calls in each set) and then used one-sample t-tests to
290 examine whether proportion of correct answers was significantly above chance per task (0.33
291 for categorization task; 0.5 for discrimination task).

292 Next, to test our hypotheses of phylogenetic distance (hypothesis 1); acoustic similarity
293 (hypothesis 2) or a combination of both (hypothesis 3), we ran GLMMs for both categorization
294 and discrimination tasks separately to examine whether species and affect predicted participant
295 accuracy expressed as the number of correct answers for each type of stimulus (species*affect
296 e.g. chimpanzee distress). We first tested the models against a null model containing only
297 intercept and random effects. All GLMMs were fitted by REML on R.studio using the
298 “bobyqa” function (optimization by quadratic approximation with a set maximum of 1’000’000
299 iterations) and the link “logit” for a standard logistic distribution of errors and a binomial
300 distribution including: Species (human, chimpanzee, bonobo, and rhesus macaque) and Affect
301 (affiliative, threat, and distress) as fixed factors, accuracy in either the discrimination or
302 categorization task as the Response Variable and participant IDs as random factor.

303 To relate our results with the acoustic analyses, we ran the same contrasts, i.e. Human vs
304 Chimpanzee, Human vs Bonobo; Chimpanzee vs Bonobo; Chimpanzee vs Macaque and
305 Bonobo vs Macaque for each affect.

306 **Results**

307 *Acoustic analyses*

308 The DA allowed us to compute Mahalanobis distances for all stimuli compared to Human
309 vocalizations (Figure 2). A GLMM analysis on Mahalanobis distances revealed the full model
310 including main effects and the interaction between Distance-Species and Affect explained
311 significantly more variance compared to the null model ($\chi^2(11) = 120.2, p < 0.001$).

312
313 Table 1: Table summarizing the statistical values for the GLMM of acoustic Mahalanobis
314 distances including main effects and the interaction.

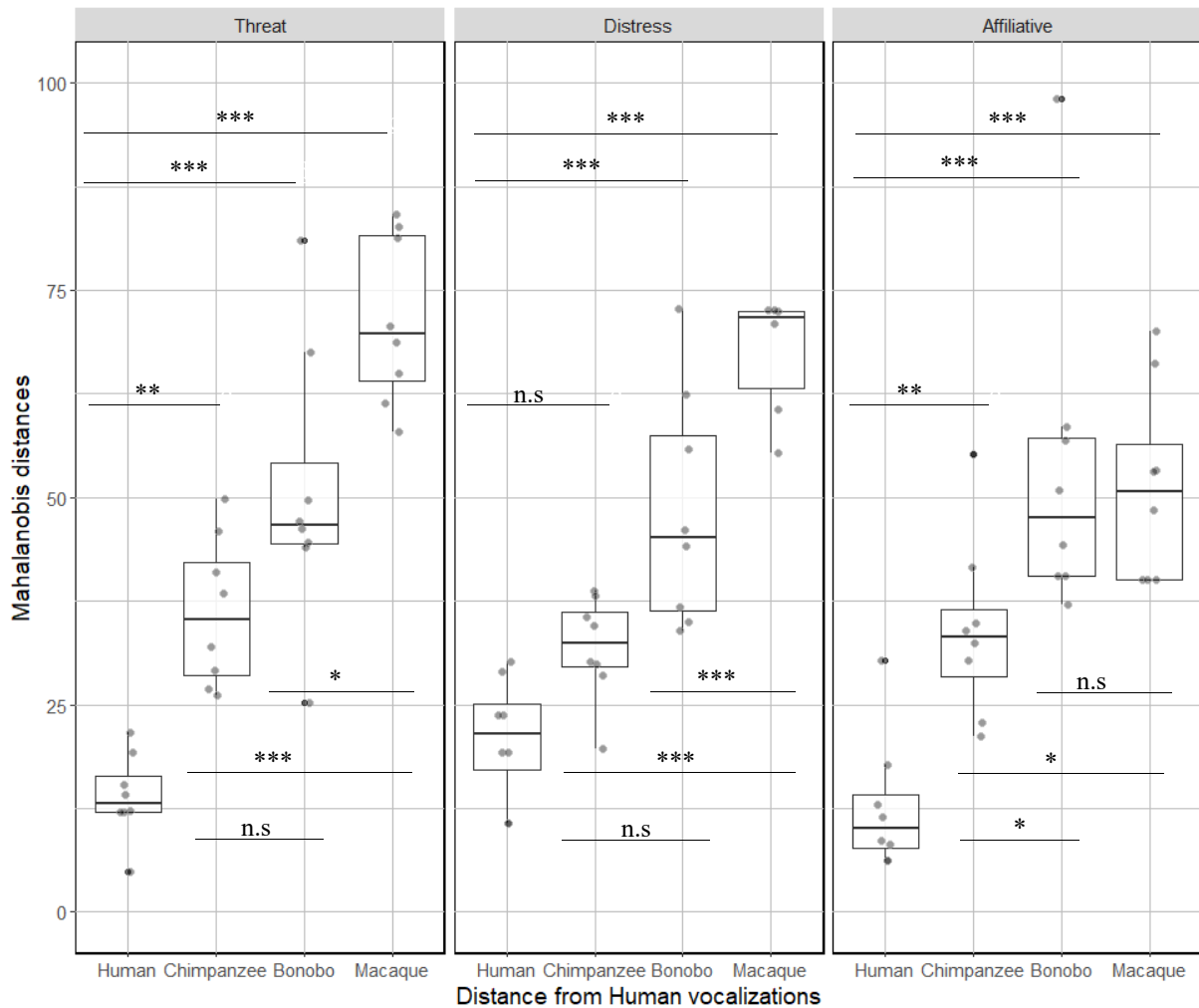
315

Summary of the model for acoustic Mahalanobis distances	Df	F value	p-value
Distance-Species	3	62.75	<.001
Affect	2	2.55	.084
Distance-Species:Affect	6	2.74	.018

316

317 The contrasts in the full model for the comparisons between the levels of distance from the
318 Human Centroid (Distance-Species) for each level of Affects are reported in Table 2 (see also
319 Figure 2). When corrected for multiple comparisons, pairwise comparisons revealed that
320 Mahalanobis distances to human centroids for human vocalisations were significantly smaller
321 than for all bonobo and all macaque vocalisations, as well as affiliative and threat chimpanzee
322 vocalizations, but not chimpanzee distress calls. Chimpanzee and bonobo vocalizations (when
323 plotted from human vocalization centroids) were not significantly different at the levels of
324 distress and threat, but bonobo affiliative vocalisations were significantly further from the
325 human centroid than chimpanzee affiliative vocalisations (see Table 2; Figure 2). Macaque
326 vocalisations were significantly further from the human centroid than chimpanzee

327 vocalisations for all affects. Macaque vocalisations were significantly further from the human
328 centroid than bonobo vocalisations for threat and distress calls, but not affiliative calls.
329



330

331 Figure 2: Boxplot of Mahalanobis distances for the 96 vocalizations representing acoustic
332 distances from human voice compared to the other species vocalizations for the different
333 affective states. Higher values represent greater acoustic distances. (* <math><0.017</math>; ** <math><0.003</math>;
334 *** <math><0.0003</math>).

335

336

337

338

339

340

341

342

343

344

345

346 Table 2: Table summarizing the results of pairwise comparisons in GLMMs for acoustic across
 347 species (Chimpanzee, Bonobo and Macaque) and affect (Threat, distress, affiliative). All p-
 348 values are compared to a corrected alpha level of 0.017 (* <0.017; **<0.003; ***<0.0003).
 349 Abbreviations: (Mac) Macaque; (Chimp) Chimpanzee; (affiliat.) affiliative.

350

	<i>Chimp threat</i>	<i>Bonobo threat</i>	<i>Mac threat</i>		<i>Chimp distress</i>	<i>Bonobo distress</i>	<i>Mac distress</i>		<i>Chimp affiliat.</i>	<i>Bonobo affiliat.</i>	<i>Mac affiliat.</i>
<i>Human threat</i>	$\chi^2(1)=10.3$; p=0.001 **	$\chi^2(1)=28.2$; p<0.001 ***	$\chi^2(1)=69.23$; p<0.001 ***	<i>Human distress</i>	$\chi^2(1)=2.58$; p=.11	$\chi^2(1)=15.8$; p<0.001 ***	$\chi^2(1)=75.72$; p<0.001 ***	<i>Human affiliat.</i>	$\chi^2(1)=9.52$; p=0.002 **	$\chi^2(1)=34.5$; p<0.001 ***	$\chi^2(1)=31.35$; p<0.001 ***
<i>Chimp threat</i>	--	$\chi^2(1)=4.42$; p=0.036	$\chi^2(1)=26.0$; p<0.001 ***	<i>Chimp distress</i>		$\chi^2(1)=5.67$; p=0.017	$\chi^2(1)=50.36$; p<0.001 ***	<i>Chimp affiliat.</i>	--	$\chi^2(1)=7.79$; p=0.005 *	$\chi^2(1)=6.32$; p=0.012 *
<i>Bonobo threat</i>		--	$\chi^2(1)=9.02$; p=0.003 *	<i>Bonobo distress</i>		--	$\chi^2(1)=22.2$; p<0.001 ***	<i>Bonobo affiliat.</i>		--	$\chi^2(1)=0.08$; p=0.78

351

352

353 Overall, while the pattern of Mahalanobis distances from the human centroid for threat
 354 vocalizations appears to mirror phylogenetic distance between species (with $H > C=B > M$),
 355 we found significant variation for both distress and affiliative vocalizations. With respect to
 356 distress calls, the pattern suggests that great ape calls are acoustically similar to each other, but
 357 different from macaque calls ($H=C=B>M$). In contrast, human affiliative calls are significantly
 358 different from all other calls, with chimpanzee calls being significantly closer to the human
 359 centroid than bonobo or macaque calls ($H>C>B=M$). The statistical analysis for all other
 360 comparisons can be found in the Supplementary Material.

361

362 *Vocal recognition performance*

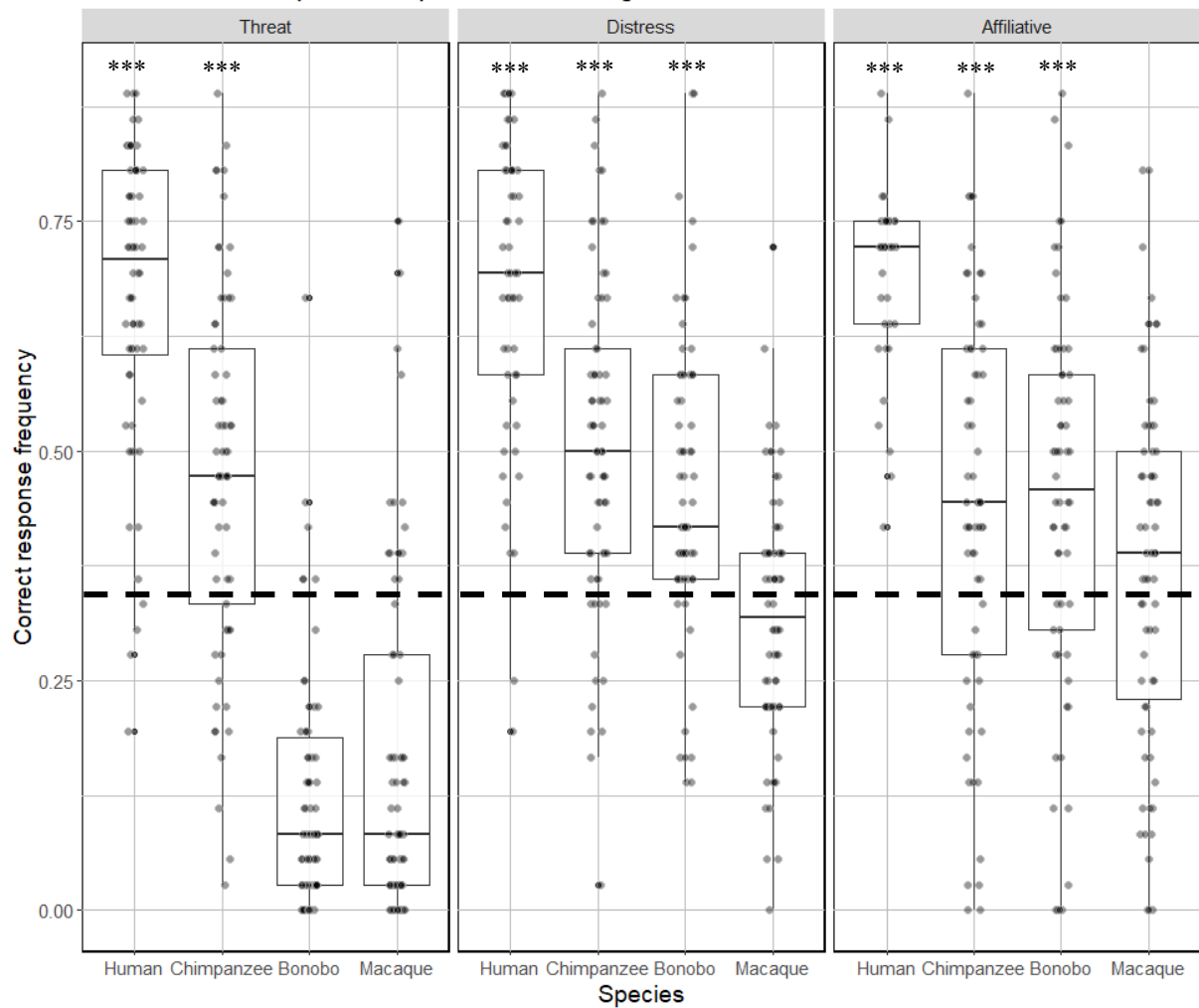
363 Patterns of performance against chance level, as well as between species and affect, differed
 364 for categorisation and discrimination.

365

366 *Categorization*

367 Participants were above chance for detecting affect for both human and chimpanzee
 368 vocalizations; this was also the case for assigning distress and affiliative calls for bonobos, but
 369 not threat calls. In contrast, no call type reached significance for macaques (Figure 3).

370



371
372

373 Figure 3: Boxplot illustrating the proportion of correct responses for each category of stimuli
374 in the Categorization task. Higher values represent greater accuracy. One sample T-test
375 analyses against chance level (0.33 - represented with the dotted line) are shown. Note that all
376 types of stimuli were categorized at a level significantly above chance, with the exception of
377 all macaque calls and threatening bonobo calls. See Table S3 in Sup Mat for the summary of
378 the t-values testing whether participants' accuracy was above chance level. *** $p < 0.001$.

379

380 A GLMM comparison for the categorization task between the null model and the full model
381 with main effects and the interaction (Species and Affects) revealed the full model explained a
382 significant amount of variance in the data $\chi^2(11) = 609.3, p < 0.001$, see Table 3).

383 Table 3: Table summarizing the main values for GLMMs of accuracy for the Categorization
384 task according to main factors and the interaction.

385

Full model for Categorization	Df	Chi-squared	p-value
Species	3	234.92	<.001
Affects	2	64.62	<.001
Species*Affects	6	17.23	<.001

386 Contrast analysis revealed that human vocalizations were systematically better recognized than
 387 chimpanzee, bonobo and macaque vocalizations across all levels of affect (Table 4). In
 388 contrast, accuracy with chimpanzee and bonobos distress and affiliative calls was similar, with
 389 chimpanzee threat calls being more accurately categorised than bonobo threat calls.
 390 Chimpanzee and bonobo distress and affiliative calls were both more accurately categorised
 391 than macaque calls. However, macaque threat calls were more accurately categorised than
 392 bonobo threat calls. All contrasts are reported in Table 4. Note that all contrasts were compared
 393 to a corrected P for multiple comparisons (Bonferroni correction: $P_{\text{corrected}} = .05/3 = .017$).

394 Table 4: Table summarizing the results of pairwise comparisons in GLMMs for categorization
 395 across species (Chimpanzee, Bonobo and Macaque) and affect (Threat, distress, affiliative. All
 396 p-values are compared to a corrected alpha level of 0.017 (* <0.017 ; ** <0.003 ; *** <0.0003).
 397 Abbreviations: (Mac) Macaque; (Chimp) Chimpanzee; (affiliat.) affiliative.
 398

	<i>Chimp threat</i>	<i>Bonobo threat</i>	<i>Mac threat</i>		<i>Chimp distress</i>	<i>Bonobo distress</i>	<i>Mac distress</i>		<i>Chimp affiliat.</i>	<i>Bonobo affiliat.</i>	<i>Mac affiliat.</i>
Human threat	$\chi^2(1)=37.37$; $p<0.001$ ***	$\chi^2(1)=304.97$; $p<0.001$ ***	$\chi^2(1)=252.77$; $p<0.001$ ***	Human distress	$\chi^2(1)=44.49$; $p<0.001$ ***	$\chi^2(1)=56.13$; $p<0.001$ ***	$\chi^2(1)=158.69$; $p<0.001$ ***	Human affiliat.	$\chi^2(1)=132.47$; $p<0.001$ ***	$\chi^2(1)=122.59$; $p<0.001$ ***	$\chi^2(1)=200.93$; $p<0.001$ ***
Chimp threat	--	$\chi^2(1)=128.84$; $p<0.001$ ***	$\chi^2(1)=95.77$; $p<0.001$ ***	Chimp distress		$\chi^2(1)=0.68$; $p=0.41$	$\chi^2(1)=35.13$; $p<0.001$ ***	Chimp affiliat.	--	$\chi^2(1)=0.19$; $p=0.66$	$\chi^2(1)=7.10$; $p<0.008$ *
Bonobo threat		--	$\chi^2(1)=2.45$; $p<0.12$	Bonobo distress		--	$\chi^2(1)=26.06$; $p<0.001$ ***	Bonobo affiliat.		--	$\chi^2(1)=9.63$; $p<0.002$ **

399

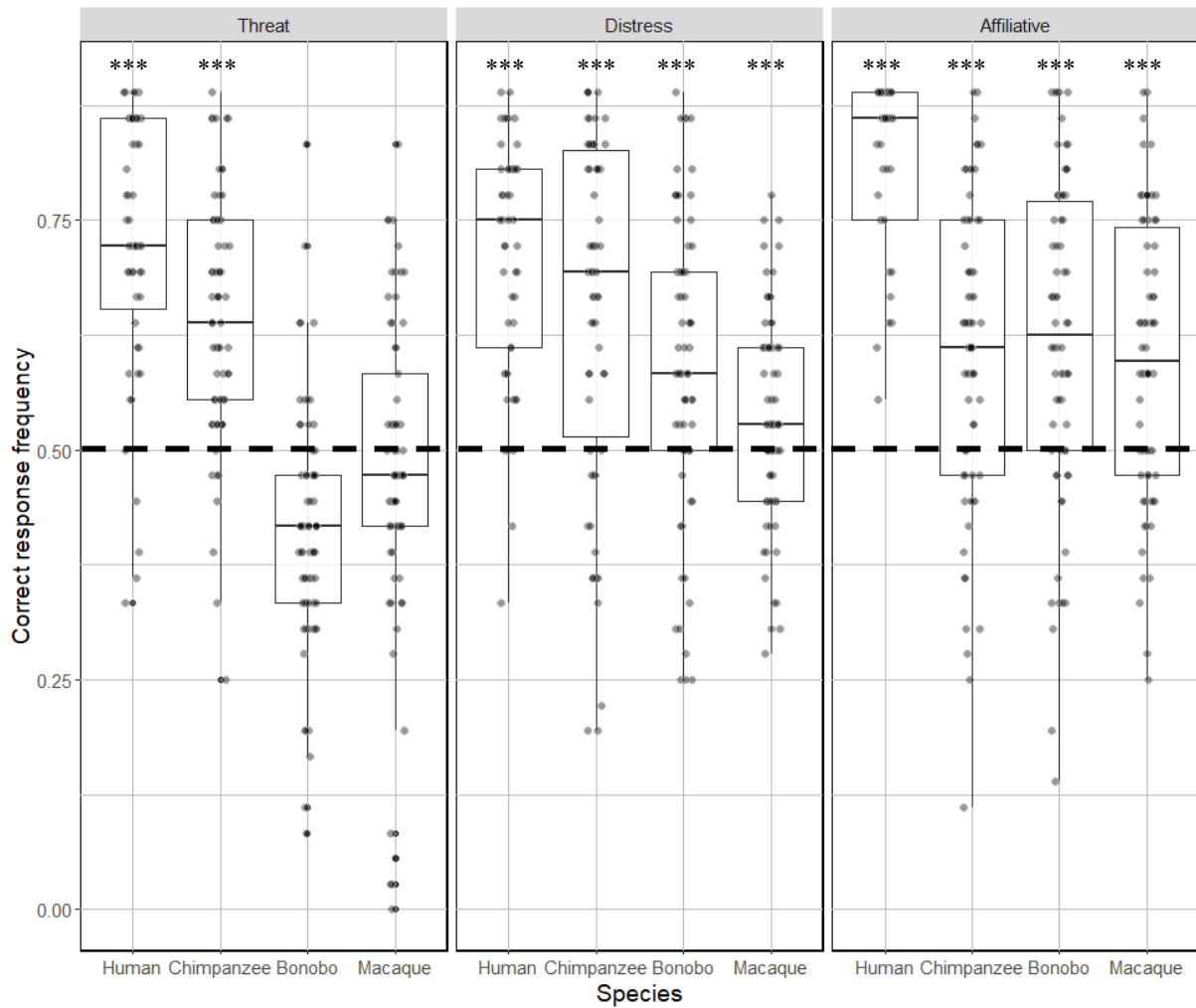
400

401 *Discrimination*

402 Participants were above chance when detecting affect for both human and chimpanzee
 403 vocalizations; this was also the case for assigning distress and affiliative calls for bonobos and
 404 macaque calls. However, threat calls for the two latter species were not discriminated at a level
 405 significantly above chance (Figure 4).
 406

407

407



408

409 Figure 4: Boxplot illustrating the proportion of correct responses in the Discrimination task.
 410 Higher values represent greater accuracy. One sample T-test analyses against chance level (0.5
 411 - shown with the dotted line) are reported. Note that all types of stimuli were discriminated at
 412 above chance levels with the exception of all macaque calls and threatening bonobo calls. ***
 413 $p < 0.001$.

414

415 A GLMM run on the discrimination task data revealed that the full model explained
 416 significantly more variation in the data than the null model $\chi^2(11) = 436.97$, $p < 0.001$, see
 417 Table 5).

418 Table 5: Table summarizing the main values for GLMMs of accuracy for the Discrimination
 419 task according to main factors and the interaction.

Full model for Discrimination	Df	Chi-squared	p-value
Species	3	150.62	<.001
Affect	2	32.52	<.001
Species*Affect	6	12.23	<.001

420 Contrast analysis revealed that human vocalizations were systematically better recognized than
 421 chimpanzee, bonobo and macaque vocalizations at all levels of affect (Table 6). Chimpanzee
 422 threat calls were significantly better discriminated compared to threat calls of both bonobo and
 423 macaques, whilst macaque threat calls were better discriminated than bonobo calls. In contrast,
 424 while participants were again significantly better at discriminating chimpanzee distress
 425 vocalizations than bonobo and macaque distress vocalizations, bonobo distress calls were
 426 discriminated better than macaque vocalizations. Finally, none of the contrasts reached
 427 significance level for comparison of affiliative vocalizations in non-human primates.

428 Table 6: Table summarizing the results of pairwise comparisons in GLMMs for discrimination
 429 across species (Chimpanzee, Bonobo and Macaque) and affect (Threat, distress, affiliative. All
 430 p-values are compared to a corrected alpha level of 0.017 (* <0.017; **<0.003; ***<0.0003).
 431 Abbreviations: (Mac) Macaque; (Chimp) Chimpanzee; (affiliat.) affiliative.

432

	<i>Chimp threat</i>	<i>Bonobo threat</i>	<i>Mac threat</i>		<i>Chimp distress</i>	<i>Bonobo distress</i>	<i>Mac distress</i>		<i>Chimp affiliat.</i>	<i>Bonobo affiliat.</i>	<i>Mac affiliat.</i>
<i>Human threat</i>	$\chi^2(1)=22.96$; p<0.001 ***	$\chi^2(1)=20.239$; p<0.001 ***	$\chi^2(1)=13.471$; p<0.001 ***	<i>Human distress</i>	$\chi^2(1)=15.57$; p<0.001 ***	$\chi^2(1)=45.77$; p<0.001 ***	$\chi^2(1)=83.47$; p<0.001 ***	<i>Human affiliat.</i>	$\chi^2(1)=12.085$; p<0.001 ***	$\chi^2(1)=11.296$; p<0.001 ***	$\chi^2(1)=12.825$; p<0.001 ***
<i>Chimp threat</i>	--	$\chi^2(1)=89.01$; p<0.001 ***	$\chi^2(1)=46.44$; p<0.001 ***	<i>Chimp distress</i>		$\chi^2(1)=7.95$; p<0.004 *	$\chi^2(1)=26.93$; p<0.001 ***	<i>Chimp affiliat.</i>	--	$\chi^2(1)=0.13$; p=0.72	$\chi^2(1)=0.11$; p=0.74
<i>Bonobo threat</i>		--	$\chi^2(1)=6.86$; p=0.009 *	<i>Bonobo distress</i>		--	$\chi^2(1)=5.62$; p=0.018	<i>Bonobo affiliat.</i>		--	$\chi^2(1)=0.49$; p=0.49

433

434 Discussion

435 In this study, we used a combination of acoustic analyses and experimental recognition tasks
 436 to investigate how humans perceive primate vocal communication of affect. Using acoustic
 437 analysis, we examined the extent to which phylogenetic proximity and the category of affect
 438 (threat, distress, affiliative) predicted call acoustic similarity in human, chimpanzee, bonobo
 439 and rhesus macaque calls. Using these acoustic analyses, we then tested whether phylogenetic
 440 similarity (hypothesis 1), acoustic distance (hypothesis 2) or a combination of both (hypothesis
 441 3) best explained human recognition of affect in these primate vocalisations. Results from two
 442 subsequent recognition tasks - discrimination and categorization - which varied on task
 443 difficulty, demonstrated that participants were generally better at categorizing and
 444 discriminating human and chimpanzee vocalizations versus bonobo and rhesus macaque calls,
 445 supporting our third hypothesis both that phylogenetic distance and acoustic similarity might
 446 influence human recognition accuracy. There was however more variation for bonobo calls,
 447 with participants having difficulty recognizing their threat calls. Finally, macaque calls were
 448 the least recognized of all primate vocalizations tested, consistent with a phylogenetic distance
 449 hypothesis.

450

451 In terms of the acoustic analyses, the acoustic factors extracted in our Discriminant Analyses
 452 revealed the crucial role of specific acoustic features such as spectral, frequency, and loudness

453 parameters (see Supp Mat) to distinguish affective vocalizations expressed by different primate
454 species. Our analysis of Mahalanobis distances showed that overall, human vocalizations in
455 the three selected affect categories were acoustically closest to chimpanzee vocalizations, with
456 distress calls virtually indistinguishable by our model. By contrast, overlap with bonobo calls
457 was much lower, despite chimpanzees and bonobos being equally phylogenetically related to
458 humans. Affiliative bonobo vocalizations also showed significant differences in acoustic
459 structure from those of chimpanzees but not from those of macaques, despite chimpanzees
460 being much more closely related to them. Note however that macaque calls were also not
461 significantly acoustically different from chimpanzee affiliative calls. The variation outlined
462 between chimpanzee and bonobo calls is in line with current evidence that despite their genetic
463 proximity, the two species have known behavioural (Gruber & Clay, 2016), neurological (Staes
464 et al., 2018) and morphological differences, including a shorter larynx for bonobos, which
465 drives a higher F0 in their vocalizations (Grawunder et al., 2018). Overall, the phylogenetic
466 hypothesis ($H < C = B < M$) was only partially supported by the distance pattern found for threat
467 vocalizations, while the rest of the affective contexts offered a mixed bag of patterns, distress
468 grouping apes together (including humans), and affiliative mostly singling out human calls.

469
470 Importantly, the acoustic similarity of chimpanzee, bonobo and rhesus monkey vocalizations
471 to those of humans did not reliably predict participants' ability to categorize and discriminate
472 their affective content. Although more accurate categorization of human vocal affect was to be
473 expected, participants were nonetheless better than chance for detecting the affective content
474 of most vocalizations of each ape species, apart from bonobo threat calls. Crucially, the latter
475 calls had been characterized as similar by the Mahalanobis analysis, suggesting that additional
476 factors come in play when recognizing primate calls. Similarly, despite the lack of acoustic
477 differences between macaque affiliative calls and other great ape affiliative vocalizations,
478 participants struggled to accurately categorize and discriminate their affective content. A
479 possibility to explain these results is that we do not know which of the acoustic factors
480 measured are the most attended to by humans; possibly skewing the weight that can be given
481 to each parameters and making their application to vocalizations that differ substantially from
482 human calls harder; further work will therefore have to fine-tune an acoustic toolbox designed
483 for human vocalisations to phylogenetically close species calls that nonetheless differ
484 acoustically from our own vocalizations. Yet, these findings should not overcast the fact that
485 our participants were generally good at classifying primate calls, particularly ape calls, with
486 the exception of bonobo threat calls. Finally, the results for rhesus macaque calls underline task
487 differences with participants above chance level for discriminating between affiliative and
488 distress calls, with the former being closest to apes' vocalizations in the Mahalanobis analysis,
489 but not the latter. This underlines once again that while acoustic distance may help participants
490 to correctly classify calls in some contexts, there may be no relation in other contexts,
491 suggesting the existence of additional factors.

492
493 Results from this study complement previous research showing highly mixed performance for
494 detecting the affective nature of rhesus monkey calls (Fritz et al., 2018; Scheumann et al., 2014;
495 Scheumann, Hasting, Zimmermann, & Kotz, 2017; Belin, Fecteau, et al., 2008) (Linnankoski
496 et al., 1994). Interestingly, our study also outlines that the differences in findings may be due

497 to the task required from the participants. Both our study and that of Linnankoski and
498 colleagues', which found some recognition of macaque affective calls, used a forced-choice
499 method (the use of two or more specific response options) to identify affective cues, whereas
500 other studies used Likert response scales. Overall, discrimination led to a higher recognition
501 for participants compared to categorization, with participants only failing to recognize threat
502 calls in bonobos and macaques. This may be due to the fact that categorization is itself more
503 complicated cognitively than discrimination (with three options rather than two), a
504 phenomenon already described when solely using human emotional calls (Dricu et al. 2017;
505 Gruber et al. 2020). Conversely, the difference between the performances in the tasks also
506 means that categorization tasks may be more discriminatory in pointing out the factors that
507 affect most the identification of the correct affect. Compared to discrimination, where patterns
508 of responses do not underline a particular hypothesis, we found that patterns of categorization
509 in the GLMM for distress and affective calls followed a phylogenetic pattern ($H > C = B > M$),
510 while the overall frequency in performance suggested an acoustic pattern for distress only
511 ($H = C = B > M$). The result for distress in particular highlights that both acoustic and phylogenetic
512 factors can be identified separately for the same affect, showing the complexity of the
513 recognition process overall; but also that categorization tasks rather than discrimination tasks
514 or Likert scales may offer the granularity necessary to identify the different intervening factors.

515

516 **Conclusion**

517 Overall, we demonstrated the ability of humans to both categorize and discriminate affective
518 cues in other primate species' vocalizations, although we found contextual differences across
519 species and affect, which are not readily explained either by phylogeny or acoustic differences.
520 Beyond single explanations, by using the acoustic distance between four primate species with
521 varying levels of phylogenetic similarity whose vocalisations also varied in different ways with
522 respect to acoustic similarity across affect categories, our study demonstrates that the
523 perception of emotional cues by humans in primate vocalizations is a complex process that
524 does not solely rely on phylogenetic or acoustic similarity. In particular, the inclusion of
525 bonobo vocalizations, while not allowing us to disentangle phylogeny from acoustic factors,
526 underlines the idiosyncratic evolutionary pathway on which they have engaged compared to
527 chimpanzees (Grawunder et al., 2018), and also suggests that there are acoustic factors partially
528 independent from phylogeny and affective content that influence the recognition of calls in
529 NHPs. In this light, bonobo calls were most often verbally pointed out by participants as the
530 most unusual. Therefore, the unfamiliarity of naïve participants with some vocalizations (e.g.
531 bonobo threatening calls) could be at play. Hence, future work will need to additionally
532 disentangle the effect of familiarity from potential acoustic parameters. It would also be
533 interesting to explore neural correlates associated with these phylogenetic and acoustic
534 parameters, to offer another level analysis to the behavioural differences outlined in the present
535 study. Finally, we hope that these new findings will contribute to a better understanding of
536 emotional processing origin in humans, by highlighting where the treatment of both primate
537 and human emotions is similar, and where our own species has differed during its evolution.

538

539

540 Acknowledgements

541 We thank Katie Slocombe very much for providing some of the auditory stimuli as well as
542 extensive comments on former versions of this preprint. We thank the Swiss National Science
543 foundation (SNSF) for supporting this interdisciplinary project (CR13II_162720 / 1 – DG-
544 TG), and the Swiss Center for Affective Sciences. ZC has received support from the ESRC-
545 ORA (ES/S015612/1), the ERC Starting Grant (802979, and CD from the foundation Ernst and
546 Lucie Schmidheiny. TG was additionally supported by a grant of the SNSF during the final re-
547 writing of this article (grant PCEFP1_186832).

548 549 References

- 550 Anderson, D. J., & Adolphs, R. (2014). A Framework for Studying Emotions Across
551 Phylogeny. *Cell*, 157(1), 187–200. <https://doi.org/10.1016/j.cell.2014.03.003>
- 552 Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., & Poeppel, D. (2015). Human
553 Screams Occupy a Privileged Niche in the Communication Soundscape. *Current*
554 *Biology : CB*, 25(15), 2051–2056. <https://doi.org/10.1016/j.cub.2015.06.043>
- 555 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
556 Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
557 <https://doi.org/10.18637/jss.v067.i01>
- 558 Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human
559 cerebral response to animal affective vocalizations. *Proceedings. Biological Sciences*,
560 275(1634), 473–481. <https://doi.org/10.1098/rspb.2007.1460>
- 561 Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A
562 validated set of nonverbal affect bursts for research on auditory affective processing.
563 *Behavior Research Methods*, 40(2), 531–539. <https://doi.org/10.3758/BRM.40.2.531>
- 564 Briefer, E. (2012). Vocal Expression of Emotions in Mammals: Mechanisms of Production and
565 Evidence. *Communication Skills*. <https://animalstudiesrepository.org/comski/1>
- 566 Brunswick, E. (1956). *Perception and the representative design of psychological experiments*.
567 University of California Press.
- 568 Davila Ross, M., Owren, M. J., & Zimmermann, E. (2009). Reconstructing the evolution of
569 laughter in great apes and humans. *Current Biology: CB*, 19(13), 1106–1111.
570 <https://doi.org/10.1016/j.cub.2009.05.028>
- 571 Dricu, M., Ceravolo, L., Grandjean, D., & Frühholz, S. (2017). Biased and unbiased perceptual
572 decision-making on vocal emotions. *Scientific Reports*, 7(1), 16274.
573 <https://doi.org/10.1038/s41598-017-16594-w>
- 574 Eyben, F., Scherer, K., Schuller, B., Sundberg, J., André, E., Busso, C., Devillers, L., Epps, J.,
575 Laukka, P., Narayanan, S., & Truong, K. P. (2016). The Geneva Minimalistic Acoustic
576 Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE*
577 *transactions on affective computing*, 7(2), 190–202.
578 <https://doi.org/10.1109/TAFFC.2015.2457417>
- 579 Ferdenzi, C., Patel, S., Mehu-Blantar, I., Khidasheli, M., Sander, D., & Delplanque, S. (2013).
580 Voice attractiveness: Influence of stimulus duration and type. *Behavior Research*

- 581 *Methods*, 45(2), 405–413. <https://doi.org/10.3758/s13428-012-0275-0>
- 582 Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support
583 categorization in very young human infants. *Proceedings of the National Academy of*
584 *Sciences*, 110(38), 15231–15235. <https://doi.org/10.1073/pnas.1221166110>
- 585 Filippi, P. (2016). Emotional and Interactional Prosody across Animal Communication
586 Systems: A Comparative Approach to the Emergence of Language. *Frontiers in*
587 *Psychology*, 7, 1393. <https://doi.org/10.3389/fpsyg.2016.01393>
- 588 Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., Hoeschele,
589 M., Ocklenburg, S., de Boer, B., Sturdy, C. B., Newen, A., & Güntürkün, O. (2017).
590 Humans recognize emotional arousal in vocalizations across all classes of terrestrial
591 vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society B:*
592 *Biological Sciences*, 284(1859), 20170990. <https://doi.org/10.1098/rspb.2017.0990>
- 593 Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and*
594 *Emotion*, 1(2), 115–143. <https://doi.org/10.1080/02699938708408043>
- 595 Frijda, N. H. (2016). The evolutionary emergence of what we call “emotions.” *Cognition and*
596 *Emotion*, 30(4), 609–620. <https://doi.org/10.1080/02699931.2016.1145106>
- 597 Fritz, T., Mueller, K., Guha, A., Gouws, A., Levita, L., Andrews, T. J., & Slocombe, K. E.
598 (2018). Human behavioural discrimination of human, chimpanzee and macaque
599 affective vocalisations is reflected by the neural response in the superior temporal
600 sulcus. *Neuropsychologia*, 111, 145–150.
601 <https://doi.org/10.1016/j.neuropsychologia.2018.01.026>
- 602 Grandjean, D., Bänziger, T., & Scherer, K. R. (2006). Intonation as an interface between
603 language and affect. In *Progress in Brain Research* (Vol. 156, pp. 235–247). Elsevier.
604 [https://doi.org/10.1016/S0079-6123\(06\)56012-1](https://doi.org/10.1016/S0079-6123(06)56012-1)
- 605 Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., &
606 Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in
607 meaningless speech. *Nature Neuroscience*, 8(2), 145–146.
608 <https://doi.org/10.1038/nn1392>
- 609 Grawunder, S., Crockford, C., Clay, Z., Kalan, A. K., Stevens, J. M. G., Stoessel, A., &
610 Hohmann, G. (2018). Higher fundamental frequency in bonobos is explained by larynx
611 morphology. *Current Biology: CB*, 28(20), R1188–R1189.
612 <https://doi.org/10.1016/j.cub.2018.09.030>
- 613 Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review*
614 *of General Psychology*, 271–299.
- 615 Gruber, T., & Clay, Z. (2016). A Comparison Between Bonobos and Chimpanzees: A Review
616 and Update. *Evolutionary Anthropology: Issues, News, and Reviews*, 25(5), 239–252.
617 <https://doi.org/10.1002/evan.21501>
- 618 Gruber, T., & Grandjean, D. M. (2017). A comparative neurological approach to emotional
619 expressions in primate vocalizations. *Neuroscience and Biobehavioral Reviews*, 73,
620 182–190.
- 621 Kamiloğlu, R. G., Slocombe, K. E., Haun, D. B. M., & Sauter, D. A. (2020). Human listeners’
622 perception of behavioural context and core affect dimensions in chimpanzee
623 vocalizations. *Proceedings of the Royal Society B: Biological Sciences*, 287(1929),
624 20201148. <https://doi.org/10.1098/rspb.2020.1148>

- 625 Kelly, T., Reby, D., Levréro, F., Keenan, S., Gustafsson, E., Koutseff, A., & Mathevon, N.
626 (2017). Adult human perception of distress in the cries of bonobo, chimpanzee, and
627 human infants. *Biological Journal of the Linnean Society*, 120(4), 919–930.
628 <https://doi.org/10.1093/biolinnean/blw016>
- 629 Kret, M. E., Prochazkova, E., Sterck, E. H. M., & Clay, Z. (2020). Emotional expressions in
630 human and non-human great apes. *Neuroscience & Biobehavioral Reviews*.
631 <https://doi.org/10.1016/j.neubiorev.2020.01.027>
- 632 Linnankoski, I., Laakso, M., Aulanko, R., & Leinonen, L. (1994). Recognition of emotions in
633 macaque vocalizations by children and adults. *Language & Communication*, 14(2),
634 183–192. [https://doi.org/10.1016/0271-5309\(94\)90012-4](https://doi.org/10.1016/0271-5309(94)90012-4)
- 635 Mahalanobis, P. C. (1936). On the Generalized Distance in statistics. In *Proceedings of*
636 *National Institute of Sciences* (Vol. 2, p. 49.55).
- 637 Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type
638 and the level of response urgency. *Proceedings of the Royal Society B: Biological*
639 *Sciences*, 268(1483), 2315–2324. <https://doi.org/10.1098/rspb.2001.1773>
- 640 Morton, E. S. (1977). On the Occurrence and Significance of Motivation-Structural Rules in
641 Some Bird and Mammal Sounds. *The American Naturalist*, 111(981), 855–869.
642 <https://doi.org/10.1086/283219>
- 643 Morton, E. S. (1982). Grading, discreteness, redundancy, and motivation-structural rules. In
644 *Acoustic Communication in Birds* (Kroodsma, D.E., Miller, E.H. and Ouellet, H., pp.
645 182–212). Academic Press.
- 646 Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, 1(3), 261–289.
647 <https://doi.org/10.1007/BF02733986>
- 648 Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., &
649 Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain
650 regions involved in processing anger prosody. *NeuroImage*, 28(4), 848–858.
651 <https://doi.org/10.1016/j.neuroimage.2005.06.023>
- 652 Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech*
653 *Communication*, 40(1–2), 227–256. [https://doi.org/10.1016/S0167-6393\(02\)00084-5](https://doi.org/10.1016/S0167-6393(02)00084-5)
- 654 Scheumann, M., Hasting, A. S., Kotz, S. A., & Zimmermann, E. (2014). The voice of emotion
655 across species: How do human listeners recognize animals' affective states? *PloS One*,
656 9(3), e91192. <https://doi.org/10.1371/journal.pone.0091192>
- 657 Scheumann, M., Hasting, A. S., Zimmermann, E., & Kotz, S. A. (2017). Human Novelty
658 Response to Emotional Animal Vocalizations: Effects of Phylogeny and Familiarity.
659 *Frontiers in Behavioral Neuroscience*, 11. <https://doi.org/10.3389/fnbeh.2017.00204>
- 660 Schore, J. R., & Schore, A. N. (2008). Modern Attachment Theory: The Central Role of Affect
661 Regulation in Development and Treatment. *Clinical Social Work Journal*, 36(1), 9–20.
662 <https://doi.org/10.1007/s10615-007-0111-7>
- 663 Staes, N., Smaers, J. B., Kunkle, A. E., Hopkins, W. D., Bradley, B. J., & Sherwood, C. C.
664 (2018). Evolutionary divergence of neuroanatomical organization and related genes in
665 chimpanzees and bonobos. *Cortex*. <https://doi.org/10.1016/j.cortex.2018.09.016>
- 666 Team, R. (2020). *RStudio: Integrated Development for R*. RStudio. RStudio, Inc.
667 <https://rstudio.com/>
- 668 Tuttle, R. H. (1993). Kano, T. 1992. The Last Ape: Pygmy Chimpanzee Behavior and Ecology.

669 Stanford University Press, Stanford, CA, xxviii + 248 pp. ISBN 0-8047-1612-9. Price
670 (hardbound), \$45.00. *Journal of Mammalogy*, 74(1), 239–240.
671 <https://doi.org/10.2307/1381928>

672

673

674