



HAL
open science

Categorization of vocal and nonvocal stimuli in Guinea baboons (*Papio papio*)

Fatima-ezzahra Ennaji, Pascal Belin, Joël Fagot

► **To cite this version:**

Fatima-ezzahra Ennaji, Pascal Belin, Joël Fagot. Categorization of vocal and nonvocal stimuli in Guinea baboons (*Papio papio*). *American Journal of Primatology*, In press, 10.1002/ajp.23387. hal-03661496

HAL Id: hal-03661496

<https://hal.science/hal-03661496v1>

Submitted on 6 May 2022

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.

1 **Categorization of vocal and non-vocal stimuli in Guinea baboons (*Papio papio*)**

2
3 Fatima-Ezzahra Ennaji^{1,2}, Joël Fagot^{2,3} and Pascal Belin^{1,3}

4 ¹ La Timone Neuroscience Institute, CNRS and Aix-Marseille University, Marseille, France

5 ² Laboratory of Cognitive Psychology, CNRS and Aix-Marseille University, Marseille, France

6 ³ Station de Primatologie-Celphedia, CNRS UAR846, 13790 Rousset, France

7
8
9 **Corresponding author**

10 Joël Fagot, Laboratoire de Psychologie Cognitive, CNRS UMR 7290 / Aix-Marseille

11 Université 3 Place Victor Hugo, Bât. 9, Case D, 13331 Marseille cedex 1.

12 E-mail : joel.fagot@univ-amu.fr

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

28 Abstract

29 Categorization of vocal sounds apart from other sounds is one of the key abilities in human
30 voice processing, but whether this ability is present in other animals, particularly non-human
31 primates, remains unclear. In the present study, 25 socially-housed Guinea baboons (*Papio*
32 *papio*) were tested on a vocal/non-vocal categorization task using Go/Nogo paradigm
33 implemented on freely accessible automated learning devices. Three individuals from the
34 group successfully learned to sort Grunt vocalizations from non-vocal sounds, and they
35 generalized to new stimuli from the two categories, indicating that some baboons have the
36 ability to develop open-ended categories in the auditory domain. Contrary to our hypothesis
37 based on the human literature, these monkeys learned the non-vocal category faster than the
38 Grunt category. Moreover, they failed to generalize their classification to new classes of
39 conspecific vocalizations (Wahoo, bark, Yak, and copulation calls), and they categorized
40 human vocalizations in the non-vocal category, suggesting that they had failed to represent
41 the task as a vocal vs non vocal categorization problem. Thus, our results do not confirm the
42 existence of a separate perceptual category for conspecific vocalizations in baboons.
43 Interestingly, the three successful baboons are the youngest of the group, with less training in
44 visual tasks, which supports previous reports of age and learning history as crucial factors in
45 auditory laboratory experiments.

46 Key words: categorization, voice perception, conspecific vocalizations, non-human primate,
47 behavior.

48

49

50

51

52

53 **Introduction**

54 Vocal perception is a key component of primate communication. Because primates use
55 vocalizations in their social environment and during social interactions, it is crucial that they
56 accurately analyze and process the available vocal information. As proposed by a
57 neurocognitive model of human voice perception (Belin et al., 2004), a basic building block
58 of vocal perception is the ability to differentiate conspecific vocalizations (CVs) from other
59 stimuli, i.e., the recognition that a given sound is emitted by the vocal tract of another
60 individual of the same species. Additional neural processes then interact to extract different
61 types of information contained in CVs, such as age, sex, identity, and affective state.
62 Preexisting cognitive abilities and neural mechanisms involved in differentiating vocal and
63 non-vocal sounds as well as extracting relevant information in CVs have likely emerged in all
64 primates, as we can extract various information even when speech is not available (e.g.,
65 identity: Latinus & Belin, 2011; gender: Mullennix et al., 1995), like non-human primates do
66 while processing CVs. Thus, a crucial question arises about voice processing, does it involve
67 specialized mechanisms not used for other non-vocal sounds from the environment and are
68 these mechanisms similar in both human and non-human primates?

69 Several neuroimaging studies revealed evidence for cortical mechanisms were activated by
70 vocal stimuli more than by other non-vocal stimuli. In humans, neural correlates of voice
71 perception were found along the superior bank of the Superior Temporal Sulcus (STS) located
72 in the auditory cortex (Belin et al., 2000, 2002). Other cortical areas were also reported to be
73 highly sensitive to vocalizations in non-human primates such as in macaque monkeys in the
74 anterior temporal lobe (Bodin et al., 2021; Petkov et al., 2008). These findings suggest the
75 existence of voice selective areas actively involved in processing and extracting information
76 carried by CVs.

77 This neural sensitivity to CVs is revealed in humans by a behavioral advantage for
78 categorizing or detecting voice stimuli, in comparison to other non-vocal sounds. For
79 instance, Agus, Suied, Thorpe and Pressnitzer (2011) found that the processing of vocal
80 stimuli was faster and more accurate in a Go/Nogo recognition task, compared to a range of
81 musical sounds. Likewise, vocal stimuli are detected fastest (as fast as 4 ms) compared to
82 instrument sounds (Agus et al., 2011; Suied et al., 2014). In one electrophysiological study,
83 Levy, Granot, Bentin (2001) reported a voice-specific response linked with greater processing
84 specificity for human voices compared with musical instruments at around 320ms after onset.
85 The above studies, and others, altogether suggest that voice stimuli are processed differently
86 from other acoustic stimuli (Charest et al., 2009; Levy et al., 2001).

87 To further investigate the behavioral mechanisms of voice processing and their evolution, we
88 asked whether non-human primates would show evidence of a separate perceptual
89 representation for CVs, as indicated by their ability to be trained to categorize CVs apart from
90 non-vocal sounds. To our knowledge, there is no behavioral evidence that monkeys possess
91 this ability. This study attempts to fill this gap and investigate a core function of voice
92 processing: categorization of CVs and non-vocal sounds.

93 Here, we chose the Guinea baboon (*Papio papio*) as a study model. Our choice of the Guinea
94 baboons is supported by the recent demonstration that their vocal productions system shares
95 anatomical (e.g., tongue anatomy) and functional traits (e.g., production of vowel-like sound;
96 Boë et al., 2017) with that of humans, and that these animals also exhibit some of the key
97 cognitive components of vocal processing. Indeed, several playback studies demonstrated
98 their capacity to extract valuable information in their CVs. Engh, Hoffmeier, Cheney and
99 Seyfarth (2006) for instance reported that female baboons responded more to the playback of
100 a dominant female's threat-grunt in case of recent threatening than when the same female had
101 recently groomed them, suggesting that these baboons have interpreted the threat-grunts based

102 on prior social interactions, and more importantly took into account the signaler's identity
103 when processing these vocalizations (Engh et al., 2006). Similarly, the response to the
104 playback of reconciliatory grunts is stronger in baboons when the receiver had a recent
105 agonistic interaction with the emitter, in comparison to controls, suggesting that they are
106 capable of distinguishing different identities based on their grunts and remembering their
107 recent aggressive interactions (Owren et al., 1997). Together, these findings make the Guinea
108 baboon an interesting model for the study of voice perception emergence, and a perfect
109 subject in a study on vocal and non-vocal categorization.

110 In the present study, we trained 25 Guinea baboons to discriminate vocal from non-vocal
111 sounds in a Go/Nogo task. First, we presented a set of baboon's vocalizations (Grunts) and
112 non-vocal auditory stimuli. We then assessed post-training performance in generalization
113 trials using novel exemplars of the two categories. Second, we presented a larger vocal
114 category using four new classes of baboon's vocalizations previously described in this species
115 (Kemp et al., 2018): barks, yaks, wahoo, and copulation calls and new exemplars of non-
116 vocal auditory stimuli. We predicted that the baboons would succeed in the first
117 categorization task and would generalize to novel auditory stimuli. Considering the ecological
118 relevance of the grunts, we expected that the baboons would express faster learning and
119 greater categorization performance for the vocal than for the non-vocal category. We also
120 predicted that the baboons would accurately categorize the four new classes of CVs in the
121 vocal category, similar to the grunt vocalization. We further inspected their performance in
122 case of human voices from both sexes at different ages (e.g., female laughter, male adult cry,
123 infant cry). Finally, we examined baboon's performance when white noise is added to the
124 vocal and non-vocal stimuli. This second experiment aimed to establish psychometric
125 functions for both categories, depending on the amount of noise added to the vocal and non-
126 vocal stimuli. With a similar procedure, Heimbauer, Beran and Owren (2021) reported that a

127 female language-trained chimpanzee had similar performance pattern to that of human
128 participants in a speech identification task, using noise-vocoded words (i.e., these stimuli
129 were made from noise bands modeled on the frequency and amplitude of natural speech).
130 They suggested that this chimpanzee might have used comparable strategies to identify
131 speech words and relied on the same cues as humans (Heimbauer et al., 2021) . If voice
132 stimuli have a special status in our study on baboons, then we expect that baboons will
133 perform better in noise for the vocal than for the non-vocal category.

134 **Methods and results**

135 **Experiment 1**

136 Experiment 1 aimed to test the ability of baboons to categorize CVs apart from non-vocal
137 sounds using the Go/Nogo task. It further investigated how baboons generalized their
138 performance when novel stimuli were introduced in the task. The first part of the experiment
139 focused on one type of CVs: the grunt. The second part explored a larger set of CVs from the
140 baboons' vocal repertoire. And the third part examined categorization performance in case of
141 human voices.

142 **Material and methods**

143 **Subjects**

144 The subjects were 25 Guinea baboons (*Papio papio*) between 3 and 24 years old. They were
145 housed in two social groups in the CNRS (Centre national de la recherche scientifique)
146 primate center CNRS, UAR846, Rousset, France. The first group comprised two adult males
147 and three adult females and the second of six males and thirteen females. Both groups were
148 housed in an outdoor enclosure connected to indoor housing used primarily at night and
149 during feeding hours. Baboons were fed daily at 5pm (fruits, vegetables, and monkey chows),

150 and water was provided ad libitum within each enclosure. Subjects were a year older during
151 the second and third parts of experiment 1.

152 **Ethics statement**

153 This research was carried out in accordance with European Union and French ethical
154 standards and received approval from the French Ministère de l'Éducation Nationale et de la
155 Recherche (approval no. APAFIS-2717-2015111708173794-V3).

156

157 **Data availability statement**

158 The data that support the findings of this study are openly available in Open Science Framework
159 depository at <http://doi.org/10.17605/OSF.IO/DZEF4>

160

161 **Apparatus**

162 The study was conducted using a total of 14 operant conditioning test systems, called
163 automated learning devices for monkeys (ALDM) which have been described in detail in
164 (Fagot & Paleressompouille, 2009). Each test system comprised a touch screen for stimulus
165 presentation, a food dispenser for reward delivery, two speakers for auditory stimulus
166 delivery, and a RFID equipment for an automatic identification of the animals once they enter
167 a test system. The larger group of baboons had free access to 10 ALDM test systems installed
168 within two trailers connected to the enclosure where the baboons live. The smaller group had
169 access to 4 ALDM test systems, installed within another trailer connected to the baboons'
170 enclosure.

171 Each time a baboon enters an ALDM test system, they are automatically identified by the
172 RFID microchips implanted in its arms. The computer then gets information on that subject

173 from a server device controlling all the ALDM test systems, to resume the task where the
174 subject left its previous trial even if on another system. This procedure allows free-self testing
175 on a 24h schedule, with no need to capture the subject for testing.

176 **General procedure**

177 The auditory task was based on a Go/Nogo response rule, with vocal sounds (Go) requiring a
178 response, and non-vocal sounds (Nogo) requiring no response; both Go, and Nogo correct
179 responses were rewarded. The Go response consisted of touching the screen where a black
180 square was presented at its center as press button. When a non-vocal stimulus was played,
181 baboons had to avoid touching the black square on the screen to be rewarded (Nogo
182 response).

183 Automatic identification of the subject triggered the trial. At the beginning of the trial, the
184 computer played a Go (vocal) or Nogo (non-vocal) auditory stimulus and simultaneously
185 displayed a black square which remained on the screen during a maximum of two seconds.
186 During that delay, the subject had to touch the square in case of Go trials or had to refrain
187 touching it otherwise. All correct (Go or Nogo) responses were rewarded by a few grains of
188 dry wheat within the ALDM system. Incorrect Go or Nogo responses (e.g., touching the
189 screen in Nogo trial or not touching it in Go trials) were followed by a 3 second time-out
190 indicated by a green screen. A minimum inter-trial interval of 3 seconds was applied between
191 two consecutive trials, but that delay could be much longer as the subject could stop and
192 resume testing at will.

193 The general timeline of the Go/Nogo task is illustrated in Figure 1.

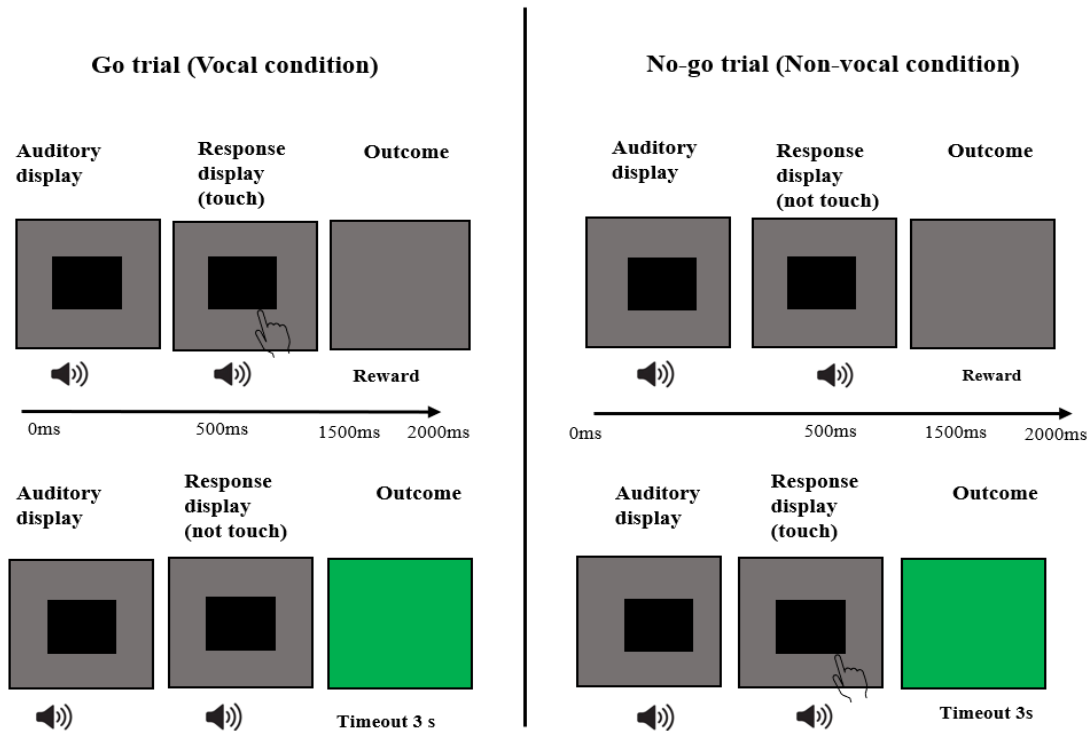


Figure 1: Illustration of the general timeline of the auditory Go/Nogo task.

194

195

196 Experiment 1.1

197 Auditory stimuli

198 Auditory stimuli were from two categories: a vocal and non-vocal category. The vocal one
 199 consisted of baboon vocalizations recorded previously from the same social group. The
 200 vocalizations chosen from their vocal repertoire in Experiment 1.1 are the “grunts,” as they
 201 are the most common affiliative vocalizations, also known to be used by all age-sex groups,
 202 usually to elicit interactions (Kemp et al., 2018). The non-vocal category consisted in various
 203 natural (e.g., water ripple, flowing stream) and environmental sounds (e.g., police noise, bells,
 204 etc.).

205 There were 95 stimuli in total, 50 stimuli for the non-vocal category and 45 stimuli for the
206 vocal one (due to smaller set of grunt recordings). Stimuli (22050Hz, mono) were manually
207 cut to a duration of 500ms and played at 65dB. White noise at an amplitude of 20% was
208 added to all stimuli to mask differences in intrinsic noise level.

209 **Training and test procedure**

210 Experiment 1.1 involved six phases which will be referred to below as P1 to P6. They are
211 described in Table 1. The rationale of our procedure was to introduce novel stimuli in each
212 phase, to assess if, and when in training, an immediate transfer of performance occurred at the
213 very first presentation of the novel stimuli. If the subjects failed to demonstrate above chance
214 performance at the first presentation of the novel stimuli, then they were continuously
215 exposed to these stimuli and previously used stimuli, until they reached learning criterion, at
216 which stage the next phase was introduced for an additional test of generalization. This
217 procedure continued from phase 1 to 6, at which point we stopped the experiment.

218 In all phases, trials were organized in blocks of 80 trials (P1, P2 and P3), 60 trials (P5) and
219 100 trials (P4 and P6) with each block containing an equal number of Go and Nogo trials.
220 Within each phase, trial blocks were repeated in a randomized order until the subject achieved
221 80% correct or more for both categories. Also, a correction procedure was implemented to
222 assist baboons during this training: in case of error, the same trial was repeated up to 3 times
223 consecutively to give the baboon a chance to respond correctly to the stimulus, before
224 presenting the next trial of the block.

225 Phase P1 consisted of presenting one single vocalization (grunt) and one non-vocal sound.
226 Phase P2 followed the same procedure as P1 but with three vocalizations (grunts) and three
227 non-vocal sounds, including those already used in P1 (i.e., 2 additional stimuli per category).
228 As described in Table 1, a progressive expansion of the stimulus set size followed in the next

229 four phases (P3, P4, P5 and P6) to reach a total of 50 stimuli in the non-vocal category and 45
 230 stimuli in the vocal one.

Phases	Number of Stimuli			
	Vocal		Non-vocal	
	Baseline	Novel	Baseline	Novel
P1	1	-	1	-
P2	1	2	1	2
P3	3	2	3	2
P4	5	5	5	5
P5	10	15	10	20
P6	25	20	30	20
Total	45		50	

231

232 **Table 1:** The table represents the generalization procedure in experiment 1.1. It shows the
 233 increasing number of novel stimuli (vocal and non-vocal) across phases (P1 to P6).

234 Results

235 We recorded a total of 300958 trials for the 25 baboons (corrected trials from the correction
 236 procedure not included). Twenty baboons failed to learn the task and never reached the
 237 learning criterion in P1, their performance is reported in supplemental materials available at
 238 <http://doi.org/10.17605/OSF.IO/DZEF4/>. Two baboons (Mako and Lomé) reached criterion in P2
 239 and P3 respectively, and three baboons (Muse, Mali and Nekke) finished the entire
 240 experiment and reached the criterion in all phases (P1 to P6). Table 2 summarizes the learning
 241 performance of these five individuals.

242

Subjects	Age (years)	Sex	Phases	Number of Blocks	Number of trials	Mean score Go	Mean score Nogo
Lomé	5	M	P1	377	30160	65.6	75
			P2	72	5760	66.6	58.5
			P3	86	6880	54.08	68.6
Mako	4	M	P1	416	33280	75.2	77.7
			P2	129	10320	70.7	66.3
Mali	4	M	P1	60	4800	73	88.8
			P2	16	1280	75	81.3
			P3	43	3440	77.4	70.4
			P4	206	20600	72.7	76.2
			P5	23	1380	62	72
			P6	24	2400	73	76
Muse	4	F	P1	6	480	78	70
			P2	2	160	77	79
			P3	1	80	85	82
			P4	28	2800	80	74
			P5	29	1740	80	69
			P6	39	3900	74	78
Nekke	3	F	P1	2	160	81	85
			P2	26	2080	74	84.6
			P3	122	9760	77.6	80.4
			P4	8	800	79	81
			P5	30	1800	78	73
			P6	15	1500	80	80

243 **Table 2:** Total number of trials and blocks in each phase for the 5 baboons, and mean score of
244 Go and Nogo trials obtained from the last 200 trials.

245 **Learning process**

246 For each successful individual (Nekke, Muse and Mali) we performed a one-way ANOVA
247 with trial phase (P1, P2, P3, P4, P5, P6) as factor using the number of blocks required to reach

248 criterion as an independent variable. For Nekke, there was a significant difference in block
 249 repetitions between phases, $F(5,197) = 30.55, p < 0.001$. Post hoc comparisons using the
 250 Tukey test revealed that the number of blocks required to reach criterion was greater in P3
 251 than in P1 ($p = 0.03$) and all remaining phases (all $ps < 0.001$). There were by contrast no
 252 reliable difference between P1, P2, P3, P4 and P6. For Muse, there was also a significant
 253 difference in block repetitions between phases, $F(5,99) = 5.13, p < 0.001$. Post hoc
 254 comparisons using the Tukey test revealed that more blocks were needed in P6 compared to
 255 P1 ($p < 0.001$), but we found no reliable difference between P1 and P5 ($p = 0.08$), P1 and P4
 256 ($p = 0.1$) and P2 and P6 ($p = 0.08$). Finally, for Mali, there was a significant difference in
 257 block repetitions between phases, $F(5,366) = 61.72, p < 0.001$. Post hoc comparisons using the
 258 Tukey test revealed that the number of block repetition was greater in P4 than in the other 5
 259 phases (all $ps < 0.001$), and there was no reliable difference between the other phases.

260 **Generalization of categorization**

261 During experiment 1.1, stimuli from both vocal and non-vocal categories were added
 262 progressively at each new phase of the experiment. To test whether subject readily
 263 generalized the categories they had formed during the previous phases to novel stimuli, we
 264 examined scores obtained at the first presentation of each novel stimulus and compared those
 265 scores to chance level (see Table 3).

266

	Go				Nogo				All trials
	P4*	P5	P6	Subtotal	P4*	P5	P6	Subtotal	Total
	(n=5)	(n=15)	(n=20)	(n=40)	(n=5)	(n=20)	(n=20)	(n=45)	(N=85)
Nekke	4 (0.8)	8 (0.53)	9 (0.45)	21 (0.52)	5 (1)	17 (0.85)	16 (0.8)	38 (0.84)	59 (0.69)

Muse	2 (0.4)	12 (0.8)	16 (0.8)	30 (0.75)	4 (0.8)	12 (0.6)	16 (0.8)	32 (0.71)	62 (0.72)
Mali	3 (0.6)	8 (0.53)	15 (0.75)	26 (0.65)	5 (1)	18 (0.9)	15 (0.75)	38 (0.84)	64 (0.75)

267 **Table 3.** Performance at the first presentation of novel stimuli. This table reports the total
 268 number of correct responses, and their corresponding proportions, obtained at the first
 269 presentation of novel Go and Nogo stimuli in phases P4, P5 and P6. Reliable above chance
 270 performance is indicated in bold and are inferred from binomial two tailed tests ($p < .05$) which
 271 were computed considering the total number of novel auditory stimuli introduced in each
 272 phase and category. * Indicate the conditions for which statistical tests were not performed
 273 due to a too small sample size.

274 Results show that all three baboons had an above chance performance in P6, in response to
 275 the presentation of novel stimuli in the non-vocal category. Two baboons (Mali and Muse)
 276 achieved an above chance performance in P6 in the vocal category and the remaining baboon
 277 (Nekke) remained below chance in the same condition. These data on generalization suggest
 278 that both Muse and Mali were able to correctly categorize novel stimuli from the two
 279 categories in P6, suggesting that they had developed open-ended representations of these
 280 categories. Results were less clear-cut for Nekke, as this subject showed an above chance
 281 category performance when the two categories are confounded, but a successful
 282 generalization only in the non-vocal category.

283 **Learning speed difference between the vocal and non-vocal categories**

284 The data presented in Table 3 suggest that the non-vocal category was learned earlier than the
 285 vocal category: the generalization of the non-vocal category (Nogo) occurred as soon as P5
 286 for Nekke and Mali. On the other hand, generalization occurred only at P6 for Muse and Mali
 287 in the vocal category (Go).

288 To explore the possibility that the subjects learned one category before the other one, we
 289 compared independently for the three individuals the number of block repetitions required to
 290 achieve 80% correct in the vocal and non-vocal categories (see Table 4). For each category, it
 291 was considered that the subjects reach the learning criterion if (1) it was above 80% correct in
 292 the considered category and (2) it was not significantly below chance, as inferred by a two-
 293 tailed binomial test, in the alternative category. This second criterion was introduced to avoid
 294 a consideration of high scores in one category that may reflect purely motor (Go or Nogo)
 295 biases, considering that such biases must be associated to an above chance performance in one
 296 category (e.g., 100% in the Go trials in case of a Go response bias) and a below chance
 297 performance in the alternative category (i.e., 0% correct in that case).

298

	Category	P1	P2	P3	P4	P5	P6	Total
Nekke	Vocal (Go)	2	26	9	8	3	9	57
	Non-vocal (Nogo)	1	1	2	1	1	3	9
Mali	Vocal (Go)	16	1	13	2	6	4	42
	Non-vocal (Nogo)	3	3	2	5	2	1	16
Muse	Vocal (Go)	1	2	1	2	1	3	10
	Non-vocal (Nogo)	2	2	1	3	2	3	13

299 **Table 4.** Number of block repetitions from P1 to P6 required to reach criterion (see main text)
 300 in the considered category. Bold character indicates when a reliable difference was obtained

301 between the two categories, as inferred by a binomial two-tailed test ($p < .05$). In that case,
302 bold characters refer to the most difficult category to learn.

303 Results show that the vocal category (Go) required more block repetitions than the non-vocal
304 category (Nogo) in three phases (P2, P3 and P4) for Nekke and in two phases (P1 and P3) for
305 Mali. For Muse, and for all the phases, the number of blocks required to reach criterion in the
306 vocal and non-vocal categories were not statistically different suggesting that these categories
307 were acquired at the same speed. However, for the two baboons Mali and Nekke, more
308 blocks were required when considering the totality of the six phases, suggesting that they have
309 learned the non-vocal category at a faster pace than the vocal one.

310 **Sensitivity and bias**

311 To measure each of the three individuals' sensitivity and examine any potential bias in their
312 responses, we calculated d' (sensitivity) based on the signal detection theory (Macmillan &
313 Creelman, 1990). Considering all trials from P6 for each subject, the "hit", "miss", "false
314 alarm", and "correct rejection" rates are determined for the vocal (Go) and non-vocal trials
315 (Nogo) and reported in Table 4. "Hit" corresponds to a correct Go response when a vocal
316 category was played, "miss" is when the subject gave a Nogo response to a vocal stimulus,
317 "false alarm" corresponds to the case where the baboon gives a Go response to a non-vocal
318 category and finally, "correct rejection" is obtained from the correct Nogo response to non-
319 vocal stimuli. d' is then determined for each individual, and for each category: $d' = z(H) -$
320 $z(F)$, where $z(H)$ and $z(F)$ are the z -transforms of "hit" and "false alarm" rates for vocal
321 category, and $d' = z(CR) - z(M)$ where $z(CR)$ and $z(M)$ are the z -transforms of "correct
322 rejection" and "miss" rates for the non-vocal category.

323 Results indicate that there was no bias in the three baboon's responses, as "hit" rates were
 324 higher than "false alarm" rates, and "correction rejection" rates were higher than "miss" rates.
 325 Also, d' values were superior to 0 for all three subjects.

Subjects	Vocal category (Go)			Non-vocal category (Nogo)		
	Hit rate (H)	False alarm rate (FA)	d'	Miss rate (M)	Correct rejection rate (CR)	d'
Nekke	0.73	0.21	1.42	0.26	0.78	1.42
Muse	0.71	0.29	1.08	0.29	0.70	1.08
Mali	0.70	0.25	1.17	0.30	0.74	1.17

326

327

328 **Table 5.** Sensitivity and response biases. This table reports the "hit", "miss", "false alarm"
 329 and "correct rejection" rates for each of the three individuals, as well as the d' .

330 **Interim discussion 1**

331 The results presented above suggest that this auditory task was extremely difficult to learn for
 332 the baboons. Regardless of the high number of trials achieved by most of the baboons, only 3
 333 successful animals from a pool of 25 were able to meet the accuracy criterion and finish all
 334 six phases of the experiment.

335 As predicted, the three successful subjects could demonstrate an open-ended form of
 336 categorical procedure, allowing to correctly categorize novel stimuli at their first presentation.
 337 Further inspection of the learning performance suggests that contrary to our predictions, the
 338 non-vocal category was learned faster and better than the vocal category, suggesting a non-

339 vocal advantage in the categorization task. These results will be discussed in greater details
340 below.

341 **Experiment 1.2**

342 After approximately 18 months, we reran Experiment 1.1 as an initial step. The same
343 procedure, apparatus and paradigm were used, with one minor modification in the paradigm:
344 we added a fixation cross on the screen where subjects had to touch to initiate the trials, and
345 new reward delivery systems (with less noise) were installed. Generalization to novel classes
346 of baboon's vocalizations (Barks, Yaks, Wahoo, and Copulation calls) and novel exemplars of
347 non-vocal stimuli (electronic, music, natural sounds etc.) was tested after retraining. We
348 predicted that the baboons would generalize to novel classes and exemplars from both
349 categories.

350 **Auditory stimuli**

351 In the vocal category, we presented four classes of baboon's vocalization. The stimuli were
352 previously recorded from the same groups as the grunt recordings of Experiment 1.1. The four
353 vocalizations chosen from their repertoire include the barks, yaks, wahoos and copulation
354 calls. Barks are vocalizations produced by both males and females at different age categories.
355 They were recorded in two contexts: prior to feeding and in the presence of sheep in the
356 surroundings. The yaks are mostly produced by infants, often coupled with a scream in case
357 of a threat or distress. The wahoos are vocalizations typically produced by adult males in
358 response to other wahoos of other individuals. Finally, copulation calls are produced by
359 females (adults, sub-adults and even juvenile) towards the end of copulation (Kemp et al.,
360 2018). In the non-vocal category, we presented a set of new various musical, environmental,
361 natural, and electronic sounds. Stimuli from both categories (22050 Hz, mono) were manually
362 cut to 500ms duration, then white noise at an amplitude of 20% was added to all stimuli to

363 mask differences in intrinsic noise level. Stimuli were played at a sound pressure level of
364 approximately 75dB(C).

365 **Procedure**

366 The individuals who were successfully retrained in the six phases (P1 to P6) of Experiment
367 1.1 were tested in Experiment 1.2. Here we used the same Go/Nogo task and general
368 procedure as described above. However, the baboons were presented with one single test
369 block of 1240 trials, corresponding to 1000 baseline trials (i.e., ten repetitions of the same 50
370 grunt trials and 50 non-vocal trials of P6 in Experiment 1.1) and 240 new trials including 120
371 non-vocal and 120 vocal trials where we presented 30 different stimuli of each of the four
372 vocalizations' classes. All new trials were presented once randomly within the 1240 trial
373 block.

374 **Results**

375 Three of the 25 baboons completed training after approximately two weeks. They were Muse,
376 Mali and Nekke, i.e., the same successful individuals as in Experiment 1.1 performed 18
377 months earlier. Three baboons (Lomé, Mako and Lips) reached P2, and the remaining
378 baboons did not even reach accuracy criterion in P1.

379 **Generalization of the categorization**

380 Table 6 reports the scores obtained at the first presentation of each novel stimulus. Results
381 show that the three subjects generalized the non-vocal category when presented with new
382 non-vocal stimuli. However, they did not generalize the new classes of CVs presented in the
383 vocal category. Moreover, the “yak” was categorized as a non-vocal sound by the three
384 subjects, as well as the “bark” by Mali and Nekke.

385

386

387

Subjects	Novel classes of CVs				Novel non-vocal sounds		
	Bark n=30	Copulation call n=30	Wahoo n=30	Yak n=30	Total n=120	Total N= 240	
Muse	12 (0.4)	16 (0.53)	12 (0.4)	7 (0.23)	47 (0.39)	78 (0.65)	125 (0.52)
Mali	9 (0.3)	19 (0.63)	16 (0.53)	9 (0.3)	53 (0.44)	79 (0.65)	131 (0.54)
Nekke	8 (0.26)	16 (0.53)	10 (0.3)	8 (0.26)	42 (0.35)	79 (0.65)	121 (0.5)

388

389 **Table 6.** Performance at the first presentation of novel stimuli. This table reports the total
 390 number of correct responses, and their corresponding proportions in brackets, obtained at the
 391 first presentation of novel vocal and non-vocal stimuli. Reliable above chance performance is
 392 indicated in bold and are inferred from binomial two tailed tests ($p < .05$) which were
 393 computed considering the total number of novel auditory stimuli introduced in each category.

394 **Sensitivity and bias**

395 Table 7 reports the d' (sensitivity) for the three subjects, considering only trials with novel
 396 stimuli for the vocal (n=120) and non-vocal category (n=120). Results show that the three
 397 baboons were less accurate in this experiment, as d' values are roughly close to 0.

398

Vocal category
(Go)

Non-vocal category
(Nogo)

Subjects	Hit rate (H)	False alarm rate (FA)	d'	Miss rate (M)	Correct rejection rate (CR)	d'
Nekke	0.35	0.34	0.02	0.65	0.65	0.02
Muse	0.39	0.35	0.11	0.60	0.65	0.11
Mali	0.43	0.34	0.24	0.56	0.65	0.24

399 **Table 7.** Sensitivity and response biases. This table reports the “hit”, “miss”, “false alarm”
400 and “correct rejection” rates for the novel stimuli for each of the three individuals, as well as
401 the d' .

402 **Interim discussion 2**

403 The result show that only the three individuals that succeeded in Experiment 1.1 were able to
404 perform again the categorization task, confirming the reliability of the findings. The results of
405 the generalization tests indicate that the barks and yaks were categorized as non-vocal,
406 whereas the copulation calls, and wahoos had better scores. The three baboons did not
407 generalize to novel classes of CVs but generalized to novel exemplars of non-vocal sounds.
408 Potential factors such as acoustic similarities between the new classes of vocalizations and the
409 grunt vocalization, and their emotional valence might explain the differences obtained among
410 the four vocal categories.

411 **Experiment 1.3**

412 The aim of this experiment was to test the baboon on the same categorization task using
413 human voices in the vocal category, and novel exemplars of non-vocal sounds in the non-
414 vocal category. We expected that the baboons would categorize human voices in the vocal
415 category and would generalize accurately to new exemplars of non-vocal stimuli.

416 **Auditory stimuli**

417 The vocal category comprised recordings from the Montreal affective voices of human voices
 418 from females and males at different ages, no specific contexts were used to classify these
 419 stimuli (e.g., female scream, enfant laughter, adult male laughter, (Belin et al., 2008). The
 420 non-vocal category was a set of new exemplars of non-vocal sounds (e.g., natural, electronic).
 421 Stimuli from both categories (22050 Hz, mono) were manually cut to 500ms duration, and
 422 played at 75dB. White noise at an amplitude of 20% was added to all stimuli to mask
 423 differences in intrinsic noise level.

424 **Procedure**

425 The same successful individuals were tested after finishing Experiment 1.2. Experiment 1.3
 426 followed the same rules as the previous one. The baboons were presented with one single
 427 block of 1240 trials, corresponding to 1000 baseline trials (same as in Experiment 1.2) and
 428 240 new trials: 120 non-vocal and 120 human vocal trials. All trials were presented randomly
 429 within the test block.

430 **Results**

431 **Generalization of the categorization**

432 Table 8 reports the scores obtained at the first presentation of each novel stimulus. None of
 433 the subjects had an above chance generalization performance for the vocal category. By
 434 contrast, Muse and Mali were above chance for the novel non-vocal sounds.

435

436

Subjects	Human voices n=120	New non-vocal sounds n=120	Total N=240
Muse	39 (0.32)	86 (0.71)	125 (0.52)

Mali	43 (0.35)	72 (0.6)	115 (0.47)
Nekke	61 (0.5)	70 (0.58)	131 (0.54)

437

438 **Table 8.** Performance at the first presentation of novel stimuli. This table reports the total
 439 number of correct responses and their corresponding proportions obtained at the first
 440 presentation of novel vocal and non-vocal stimuli. Reliable above chance performance is
 441 indicated in bold and are inferred from binomial two tailed tests ($p < .05$) which were
 442 computed considering the total number of novel auditory stimuli introduced in each category.

443 Sensitivity and bias

444 We calculated d' (sensitivity) for the three subjects considering trials with novel stimuli
 445 ($n=120$) for the vocal category and ($n=120$) for the non-vocal one. Results show that again,
 446 the three baboons were less accurate in this experiment (d' close to 0). Moreover, Mali had a
 447 negative d' suggesting that this baboon responded below chance level.

Subjects	Vocal category (Go)		d'	Non-vocal category (Nogo)		d'
	Hit rate (H)	False alarm rate (FA)		Miss rate (M)	Correct rejection rate (CR)	
Nekke	0.5	0.41	0.23	0.49	0.58	0.23
Muse	0.32	0.28	0.12	0.67	0.71	0.12
Mali	0.35	0.4	-0.13	0.65	0.6	-0.13

448 **Table 9.** Sensitivity and response biases. This table reports the “hit”, “miss”, “false alarm”
 449 and “correct rejection” rates for each of the three individuals, as well as the d' .

450 Interim discussion 3

451 Contrary to our predictions, the results reveal that human voices were not categorized as vocal
 452 sounds. Two individuals even categorized them as non-vocal sounds. The non-vocal sounds in

453 the other hand, were accurately categorized, similar to the previous experiments. Also, d'
454 results suggest that the individuals gave Nogo responses more often whenever they were
455 uncertain about the presented stimuli. This hypothesis will be discussed below.

456 **Experiment 2**

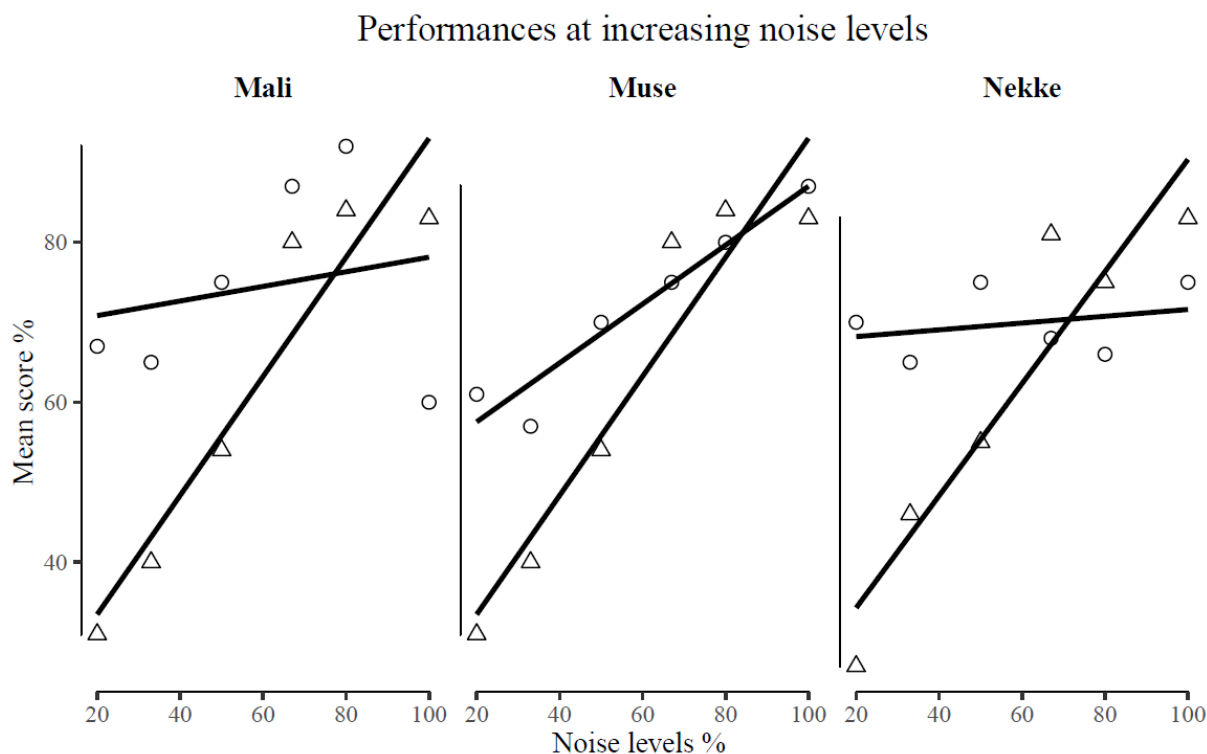
457 The aim of this experiment was to test the three successful baboons on their ability to
458 discriminate vocal and non-vocal sounds under different levels of added noise to each
459 category and examine whether one of the two categories would be more (or less) affected by
460 the noise.

461 **Procedure**

462 Experiment 2 used the same Go/Nogo task and general procedure, as described above. We
463 used stimuli from Experiment 1.1, with grunt vocalizations for the vocal category and non-
464 vocal sounds for the non-vocal one. We added increasing levels of white noise to the stimuli,
465 ranging from 20% to 100% (complete noise) in 6 levels (20%, 33%, 50%, 66%, 80% and
466 100%). The baboons were presented with one single block of 2000 trials, corresponding to
467 1000 trials for each vocal and non-vocal condition, which were randomized and presented
468 once and without the correction procedure. Baboons had to follow the same Go/Nogo
469 response rule as for previous experiments regardless of the noise level.

470 **Results**

471 Figure 2 illustrates the variations of performance with noise level, for the vocal and non-vocal
472 categories. It shows a similar pattern for all three subjects, with a performance decrease that
473 especially affected the vocal category (Go) as noise levels increased, whereas performances in
474 the non-vocal category (Nogo) remained rather high regardless of noise levels.



475

476 **Figure 2.** Percentage of Nogo response for the three baboons in the vocal (triangle) and non-
 477 vocal (circle) categories, depending on noise level (20%, 33%, 50%, 67%, 80%, 100%). Note
 478 that the Nogo responses corresponded to correct responses for the non-vocal (Nogo) trials, but
 479 to an error response for the vocal (Go) trials.

480 To examine the effect of noise level (20%, 33%, 50%, 67%, 80%, 100%) and condition
 481 (vocal, non-vocal) on performance, we performed a two-way ANOVA on the mean scores for
 482 each subject. For Nekke, results showed a significant main effect of condition (vocal, non-
 483 vocal), $F(1,2011) = 7.16, p = 0.007$, a significant main effect of noise levels, $F(5,2011) =$
 484 $13.05, p < 0.001$, and a significant interaction between noise levels and condition, $F(5,2011) =$
 485 $17.39, p < 0.001$. Post hoc Tukey tests revealed a significant difference between the vocal and
 486 non-vocal conditions at 100% ($p < 0.001$), 80% ($p < 0.001$), and at 67% noise level ($p < 0.001$).
 487 For Muse, results showed a significant main effect of condition, $F(1,2003) = 4.52, p = 0.03$, a
 488 significant main effect of noise levels, $F(5,2003) = 5.25, p < 0.001$, and a significant
 489 interaction between noise levels and condition $F(5,2003) = 25.92, p < 0.001$. Post hoc Tukey

490 tests revealed significant differences between the vocal and non-vocal conditions at noise
491 levels 100% ($p < 0.001$), 80% ($p < 0.001$), 67% ($p < 0.001$), and 20% ($p < 0.001$).

492 Finally, for Mali, results showed a significant main effect of condition, $F(1,1992) = 17.82$, p
493 < 0.001 , a significant main effect of noise levels, $F(5,1992) = 7.2$, $p < 0.001$, and a significant
494 interaction between noise levels and condition, $F(5,1992) = 23.52$, $p < 0.001$. Post hoc Tukey
495 test revealed a significant difference between the vocal and non-vocal conditions at noise
496 levels 100% ($p < 0.001$), 80% ($p < 0.001$), 67% ($p < 0.001$), and 50% ($p = 0.004$).

497 **Interim discussion 4**

498 A similar pattern was observed for the three baboons. As noise levels increased, more Nogo
499 responses associated to the non-vocal category were performed. Within the first noise levels
500 and until 50%, the baboons performed the categorization task accurately, suggesting that they
501 were still able to respond based on the already learned task following the same rules.
502 However, as noise levels continued to increase, the baboons considered all auditory stimuli as
503 non-vocal, because of the absence of enough vocal information that allows them to
504 differentiate the two categories. In line with the findings of Experiment 1.3, baboons gave a
505 Nogo response when the auditory stimuli were perceived as different from the original grunts,
506 due to the addition of noise to these grunts. We predicted that one of the two categories would
507 be affected by the added noise, according to the three baboon's performance, categorization of
508 the vocal stimuli was more affected than the non-vocal stimuli. This result is in accordance
509 with the idea that the vocal category (here restricted to grunts) is a more narrow – less
510 inclusive - category than the non-vocal one. The addition of a large amount of noise (i.e.,
511 $> 50\%$) to the grunt stimuli prevents an accurate classification of these stimuli in the vocal
512 (i.e., grunt) category, leading to a larger amount of non-vocal (i.e., Nogo response) responses.

513

514 **Discussion**

515 The aim of the present study is to investigate the ability of baboons to categorize vocal and
516 non-vocal stimuli. Two social groups of free-ranging baboons were tested in a computerized
517 task, using a Go/Nogo paradigm. Throughout all experiments (see methods and results), three
518 individuals out of the 25 baboons tested were able to perform the tasks. We will discuss the
519 results in detail below.

520 **Category learning and categorical performance:**

521 In Experiment 1, we expected that the baboons would learn to categorize grunt vocalizations
522 and non-vocal sounds and generalize to novel stimuli from both categories. Three baboons
523 (Mali, Muse, and Nekke) succeeded all six phases of generalization in Experiment 1.1, two
524 others succeeded the initial phases (Mako and Lomé reached P2 and P3 respectively), and the
525 remaining 20 baboons did not achieve criterion to pass P1. They were not lacking the
526 motivation to perform the task as reflected in the important number of trials performed.
527 Inspection of Table 2 shows that the three baboons learned the task at a different pace. Nekke
528 was for instance remarkably fast to learn at P1 (only two blocks), while Muse required a little
529 bit more block (n=6) and Mali much more (n=60). Moreover, these subjects expressed no
530 consistent trend in the number of blocks required to reach criterion when they were exposed
531 to an increasing number of auditory stimuli during the six training phases. For instance, Mali
532 and Nekke required a numerically greater number of training blocks in P1 than in P6 (Mali:
533 P1=60, P6=24; Nekke: P1=2, P6=15), but the opposite pattern was found in Muse (P1=6,
534 P6=39). We note however that two subjects indicated a sharp increase in the number of blocs
535 required to reach criterion in either P3 (Muse: 122 blocks) or P4 (Mali: 206 blocks).

536 Learning new categories in both humans and other animals generally proceeds from an initial
537 form of rote learning to a more flexible open-ended form of categorization allowing

538 generalization to novel stimuli (Herrnstein, 1990). We assume that the peaks reported above
539 in P3 (Muse) and P4 (Mali) correspond to the period when the strategy relying on rote
540 learning became inefficient to solve the task with a too large number of stimuli per class
541 ($n=10$ in P3 and $n=20$ in P4), calling for the development of more advanced (open-ended)
542 categorical representations of the two classes. Open-ended categorization is demonstrated
543 when the categorical responses generalize at the first presentation of novel exemplars of the
544 classes. Table 3 confirms that Mali and Muse reached this level of performance later in P6.
545 The results were less clear-cut for Nekke. This subject demonstrated an above chance
546 generalization performance for the non-vocal category only (P5 and P6), suggesting that
547 Nekke achieved a narrower form of categorization compared to Muse and Mali. Altogether,
548 our results suggest that 3 out of 25 baboons learned to categorize grunts and non-vocal
549 sounds, and two generalized to novel stimuli from both categories.

550 **Do baboons demonstrate a vocal advantage?**

551 We predicted that the baboons would learn faster and more efficiently the vocal category
552 (e.g., grunts) as these calls have more ecological relevance than the non-vocal sounds.
553 Contrary to our expectations, Table 4 shows that the non-vocal category was acquired earlier
554 during training than the vocal one for Nekke and Muse and required less training blocks to
555 reach training criterion in P2, P3 and P4, and in P1 and P3 for Nekke and Mali, respectively
556 (Table 4). These results are at odd with studies on humans that shows that participants can
557 more easily categorize stimuli from the vocal category.

558 For instance, in Bigand, Delbé, Gérard and Tillmann (2011), human participants were asked
559 to classify stimuli in three main categories: Human voice, musical and environmental sounds
560 at increasing durations. They found that voices were the most easily recognizable, hence
561 higher accuracy for voices than for the two other classes (Bigand et al., 2011). Similarly,
562 Suied, Agus, Thorpe, Mesgarani and Pressnitzer (2014) asked human listeners to indicate

563 whether a sound was part of the voice category or the distractor category in a yes/no
564 recognition task. They were able to detect voice stimuli from the other sound categories at
565 very brief durations of 4ms. In the other hand, instruments were recognized at duration
566 starting at 8ms (Suied et al., 2014). The advantage for the non-vocal category obtained in
567 baboons could be explained by the nature of our two categories of stimuli. In Experiment 1.1,
568 the vocal category remained a narrow category limited to grunt vocalizations, while the non-
569 vocal category was much more heterogeneous (composed of environmental, natural, and
570 electronic sounds). This asymmetry between the two categories could have favored a faster
571 opening of the non-vocal category during training, in comparison to the vocal one. This
572 hypothesis was examined in Experiment 1.2 and discussed next.

573 **Categorization of new auditory stimuli: Novel CVs' classes and human voices**

574 Before experiment 1.2, we exposed again our baboons to P1 to P6 as a preliminary phase after
575 18 months break. This replication confirmed the reliability of our findings, as the same three
576 individuals succeeded once again, while the remaining individuals failed again. Presentation
577 of new classes of conspecific vocalizations (Wahoo, bark, yak, and copulation call) and new
578 non vocal sounds in Experiment 1.2 confirmed that the non-vocal category was the only open-
579 ended category allowing generalization to new stimuli (see Table 6). The lack of
580 generalization to novel vocal sounds suggests that these baboons have developed in
581 Experiment 1.1 a strategy involving two categories, the grunts and only the grunts on the one
582 hand, and the other sounds on the other hand, and then kept this strategy when they were
583 presented with novel vocal stimuli in Experiment 1.2. In other words, they solved the tasks by
584 touching the screen (Go-response) in case of grunts and not touching the screen (Nogo) in
585 case of “non-grunt.” This strategy was further confirmed in Experiment 1.3 where we
586 presented human voices in the vocal category and added new non-vocal sounds in the non-
587 vocal one. Consistent with Experiments 1.1 and 1.2, none of the baboons categorized human

588 voices as vocal sounds and Muse and Mali categorized reliably human voices in the non-vocal
589 category (Table 8).

590 Results of Experiment 2 further support the conclusion that the baboons relied on a strategy
591 distinguishing the grunts from the non-grunt stimuli. This experiment revealed a similar
592 pattern in our three baboons: their performance in the non-vocal category did not considerably
593 vary as noise levels increased, but their performance in the vocal category decreased to
594 chance level (as early as 50% of noise) or below when the noise level increased (Figure 2).
595 These results suggest that the vocal category did not resist the added white noise. When
596 baboons could not process enough features to detect the grunt vocalizations, because of the
597 white noise, they categorized them as “non-grunt” and gave a Nogo response, in accordance
598 with grunt vs non-grunt categorization strategy.

599 **Age and past-experience factors**

600 We note that Nekke, Muse and Mali are the youngest subjects of our group (3-years old for
601 Nekke, 4 for Muse and Mali), suggesting a possible effect of age in our study. In support of
602 this idea, 2 other young baboons were successful in the first phases of Experiment 1.1, Lomé
603 (5 years old) passed criterion from P1 to P3 but did not achieve criterion to P4, and Mako (4
604 years old) passed criterion from P1 to P2 (see Table 2). Similar results were reported in other
605 studies involving monkeys. For instance, Amato & Colombo (1986) could only successfully
606 train 4 out of 8 Tufted capuchin monkeys (*Cebus apella*) in an auditory matching-to-sample
607 task. These four monkeys were aged between 4 and 6 years old, whereas three of the
608 remaining monkeys were considerably older (14, 21 and 22 years old). They suggested that
609 age might be a factor in auditory experiments. This age effect could lie in perception. In
610 humans, the younger individuals have a better auditory acuity than older individuals, who
611 suffer from audiometric loss indicated by thresholds elevation (Humes & Roberts, 1990) and
612 general slowing in auditory processing (Alain et al., 2012). Although not tested, these effects

613 of age can also derive from variation in training. Thus, D'Amato & Colombo (1986) proposed
614 that past exposure to visual tasks can also affect training performance in auditory tasks. In
615 their study, they reported that the first experimental training had been with acoustic stimuli for
616 the successful monkeys, in sharp contrast with the less successful monkeys who had previous
617 exposures to visual stimuli. Accordingly, the group of baboons tested in the present study had
618 significant training with complex tasks in the visual modality. Thus, it is perhaps not
619 surprising that the youngest baboons that had less visual training in the past were the
620 successful ones. Overall, our results suggest also that the amount of previous experience with
621 visual/auditory modalities, and age might be crucial in auditory tasks.

622 **General conclusion**

623 We aimed to test the ability of Guinea baboons to discriminate vocal from non-vocal sounds,
624 which is a core ability of vocal processing. Several playback studies have shown that Guinea
625 baboons and other non-human primate's species can process and extract different information
626 conveyed in their conspecific vocalizations (e.g., identity: Engh et al., 2006; Miller & Wren
627 Thomas, 2012; Rendall et al., 1998; Snowdon & Cleveland, 1980, familiarity and kinship:
628 Cheney & Seyfarth, 1999; Pfefferle et al., 2013, 2016), presence and nature of predators:
629 Cheney & Seyfarth, 1980; Fischer et al., 1995, 2001, 2002; Kirchof & Hammerschmidt,
630 2006). Until now, there is to our knowledge no behavioral evidence that they can differentiate
631 CVs from other sounds (natural, environmental, heterospecific vocalizations etc.). In the
632 current study, three young baboons learned to categorize grunts vocalization and non-vocal
633 sounds. The current data suggest that subjects learned to categorize grunt and "non-grunt"
634 sounds but failed to represent the task as a vocal vs non-vocal categorization problem.
635 Because they did not generalize to other classes of baboon's vocalizations and categorized
636 human voices as non-vocal sounds, we are unable to draw any definitive conclusion on their
637 ability to process voice cue as such, and the evolution of this ability. By contrast, the accurate

638 classification of novel vocal and non-vocal stimuli in Experiment 1.1 clearly demonstrates
639 that the baboons have the ability to develop open-ended categories in the auditory domain.

640 In a different perspective, the fact that only three baboons could learn to categorize auditory
641 stimuli confirms that addressing questions on auditory processing in non-human primates is
642 very challenging in laboratories. Similar difficulties were already encountered in past studies
643 from other research groups. For instance, D'Amato & Colombo (1986) tested 8 capuchin
644 monkeys (*Cebus apella*) in a matching-to-sample task and found that the few (n=4) successful
645 monkeys were the youngest ones, and those who had only been trained with acoustic stimuli.

646 In another study, Stepien & Cordeau (1960) trained vervet monkeys (*Cercopithecus aethiops*,
647 n=7) to discriminate two auditory stimuli (pure tones), and reported that the monkeys learned
648 the task but only after using a simplified version of the task with more positively reinforced
649 trials. They also reported that all the monkeys were experimentally naïve. Nonetheless, the
650 training was particularly difficult and lasted longer than the visual version of the same task
651 (Stepien & Cordeau, 1960). Similarly, Kojima (1985) tested two Japanese monkeys on a
652 Go/Nogo delayed matching to sample, and reported that the acquisition of auditory matching
653 was more difficult, and performances were more fragile compared to the visual version of the
654 task (Kojima, 1985). Therefore, our findings do not depart from the literature in this respect.

655 The source of such difficulties remains unclear at this point, but evidence suggest that it has a
656 multifactorial origin: the performance of the individual seems to depend on the
657 appropriateness of the training and test protocol, the age of the subjects, and their past
658 experience with visual and auditory discrimination problems.

659 **Acknowledgements**

660 This work was supported by the European Research Council (ERC) under the European
661 Union's Horizon 2020 research and innovation program to Pascal Belin. This research was

662 conducted at Station de Primatologie-Celphedia, CNRS UAR846, Rousset, France. The
663 authors thank its staff for technical support.

664

665 **References**

666 Agus, T. R., Suied, C., Thorpe, S. J., Pressnitzer, D., Agus, T. R., Thorpe, S. J., & Pressnitzer,
667 D. (2011). *Characteristics of human voice processing. May 2010*, 509–512.

668 Alain, C., McDonald, K., & Van Roon, P. (2012). Effects of age and background noise on
669 processing a mistuned harmonic in an otherwise periodic complex sound. *Hearing*
670 *Research*, 283(1–2), 126–135. <https://doi.org/10.1016/j.heares.2011.10.007>

671 Amato, M. R. D., & Colombo, M. (1986). *Auditory matching-to-sample in monkeys (Cebus*
672 *apella). 13(4)*, 375–382.

673 Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice
674 perception. *Trends in Cognitive Sciences*, 8(3), 129–135.
675 <https://doi.org/10.1016/j.tics.2004.01.008>

676 Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A
677 validated set of nonverbal affect bursts for research on auditory affective processing.
678 *Behavior Research Methods*, 40(2), 531–539. <https://doi.org/10.3758/BRM.40.2.531>

679 Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds.
680 *Cognitive Brain Research*, 13(1), 17–26. [https://doi.org/10.1016/S0926-6410\(01\)00084-](https://doi.org/10.1016/S0926-6410(01)00084-2)
681 2

682 Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in
683 human auditory cortex. *Nature*, 403, 309–312.

- 684 Bigand, E., Delbé, C., Gérard, Y., & Tillmann, B. (2011). Categorization of extremely brief
685 auditory stimuli: Domain-specific or domain-general processes? *PLoS ONE*, 6(10).
686 <https://doi.org/10.1371/journal.pone.0027024>
- 687 Bodin, C., Trapeau, R., Nazarian, B., Sein, J., Degiovanni, X., Baurberg, J., Rapha, E.,
688 Renaud, L., Giordano, B. L., & Belin, P. (2021). Functionally homologous
689 representation of vocalizations in the auditory cortex of humans and macaques. *Current*
690 *Biology*, 1–6. <https://doi.org/10.1016/j.cub.2021.08.043>
- 691 Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., Becker, Y.,
692 Rey, A., & Fagot, J. (2017). Evidence of a Vocalic Proto-System in the Baboon (*Papio*
693 *papio*) Suggests Pre-Hominin Speech Precursors. *Plos One*, 12(1), e0169321.
694 <https://doi.org/10.1371/journal.pone.0169321>
- 695 Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S.,
696 Chartrand, J. P., & Belin, P. (2009). Electrophysiological evidence for an early
697 processing of human voices. *BMC Neuroscience*, 10, 127. [https://doi.org/10.1186/1471-](https://doi.org/10.1186/1471-2202-10-127)
698 [2202-10-127](https://doi.org/10.1186/1471-2202-10-127)
- 699 Cheney, D. L., & Seyfarth, R. M. (1980). Vocal recognition in free-ranging vervet monkeys.
700 *Animal Behaviour*, 28(2), 362–367. [https://doi.org/10.1016/S0003-3472\(80\)80044-3](https://doi.org/10.1016/S0003-3472(80)80044-3)
- 701 Cheney, Dorothy L., & Seyfarth, R. M. (1999). *Recognition of other individuals' social*
702 *relationships by female baboons*. 67–75.
- 703 Engh, A. L., Hoffmeier, R. R., Cheney, D. L., & Seyfarth, R. M. (2006). Who, me? Can
704 baboons infer the target of vocalizations? *Animal Behaviour*, 71(2), 381–387.
705 <https://doi.org/10.1016/j.anbehav.2005.05.009>
- 706 Fagot, J., & Paleressompouille, D. (2009). Automatic testing of cognitive performance in

- 707 baboons maintained in social groups. *Behavior Research Methods*, 41(2), 396–404.
708 <https://doi.org/10.3758/BRM.41.2.396>
- 709 Fischer, J., Hammerschmidt, K., Cheney, D. L., & Seyfarth, R. M. (2002). Acoustic features
710 of male baboon loud calls: Influences of context, age, and individuality. *The Journal of*
711 *the Acoustical Society of America*, 111(3), 1465–1474.
712 <https://doi.org/10.1121/1.1433807>
- 713 Fischer, J., Hammerschmidt, K., & Todt, D. (1995). Factors Affecting Acoustic Variation in
714 Barbary- macaque (*Macaca sylvanus*) Disturbance Calls. *Ethology*, 101(1), 51–66.
715 <https://doi.org/10.1111/j.1439-0310.1995.tb00345.x>
- 716 Fischer, J., Metz, M., Cheney, D. L., & Seyfarth, R. M. (2001). Baboon responses to graded
717 bark variants. *Animal Behaviour*, 61(5), 925–931.
718 <https://doi.org/10.1006/anbe.2000.1687>
- 719 Heimbauer, L. A., Beran, M. J., & Owren, M. J. (2021). A chimpanzee recognizes varied
720 acoustical versions of sine-wave and noise-vocoded speech. *Animal Cognition*, 24(4),
721 843–854. <https://doi.org/10.1007/s10071-021-01478-4>
- 722 Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition*, 37(1–
723 2), 133–166. [https://doi.org/10.1016/0010-0277\(90\)90021-B](https://doi.org/10.1016/0010-0277(90)90021-B)
- 724 Humes, L. E., & Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired
725 elderly: The contributions of audibility. *Journal of Speech and Hearing Research*, 33(4),
726 726–735. <https://doi.org/10.1044/jshr.3304.726>
- 727 Kemp, C., Rey, A., Legou, T., Boë, L.-J., Berthommier, F., Becker, Y., & Fagot, J. (2018).
728 Vocal repertoire of captive Guinea baboons (*Papio papio*). In *Origins of Human*
729 *Language: Continuities and Discontinuities with Nonhuman Primates*.

730 <https://doi.org/10.3726/b12405>

731 Kirchhof, J., & Hammerschmidt, K. (2006). Functionally referential alarm calls in tamarins
732 (*Saguinus fuscicollis* and *Saguinus mystax*) - Evidence from playback experiments.

733 *Ethology*, 112(4), 346–354. <https://doi.org/10.1111/j.1439-0310.2006.01165.x>

734 Kojima, S. (1985). Auditory short-term memory in the Japanese monkey. *International
735 Journal of Neuroscience*, 25(3–4), 255–262.

736 <https://doi.org/10.3109/00207458508985378>

737 Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, 21(4), 1–3.

738 <https://doi.org/10.1016/j.cub.2010.12.033>

739 Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli:
740 Electrophysiological evidence. *NeuroReport*, 12(12), 2653–2657.

741 <https://doi.org/10.1097/00001756-200108280-00013>

742 Macmillan, N. A., & Creelman, C. D. (1990). Response Bias: Characteristics of Detection
743 Theory, Threshold Theory, and “Nonparametric” Indexes. *Psychological Bulletin*,

744 107(3), 401–413. <https://doi.org/10.1037/0033-2909.107.3.401>

745 Miller, C. T., & Wren Thomas, A. (2012). Individual recognition during bouts of antiphonal
746 calling in common marmosets. *Journal of Comparative Physiology A: Neuroethology,*

747 *Sensory, Neural, and Behavioral Physiology*, 198(5), 337–346.

748 <https://doi.org/10.1007/s00359-012-0712-7>

749 Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The
750 Perceptual Representation of voice gender. *Acoustic Society of America*, 98(6), 23–57.

751 https://doi.org/10.1007/978-3-030-14832-4_2

752 Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like

- 753 grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production
754 processes and functions . *The Journal of the Acoustical Society of America*, *101*(5),
755 2951–2963. <https://doi.org/10.1121/1.418523>
- 756 Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K.
757 (2008). A voice region in the monkey brain. *Nature Neuroscience*, *11*(3), 367–374.
758 <https://doi.org/10.1038/nn2043>
- 759 Pfefferle, D., Hammerschmidt, K., Mundry, R., Ruiz-Lambides, A. V., Fischer, J., & Widdig,
760 A. (2016). Does the structure of female Rhesus Macaque coo calls reflect relatedness
761 and/or familiarity? *PLoS ONE*, *11*(8), 1–16.
762 <https://doi.org/10.1371/journal.pone.0161133>
- 763 Pfefferle, D., Ruiz-Lambides, A. V., & Widdig, A. (2013). Female rhesus macaques
764 discriminate unfamiliar paternal sisters in playback experiments: Support for acoustic
765 phenotype matching. *Proceedings of the Royal Society B: Biological Sciences*,
766 *281*(1774). <https://doi.org/10.1098/rspb.2013.1628>
- 767 Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in
768 identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations . *The Journal of the*
769 *Acoustical Society of America*, *103*(1), 602–614. <https://doi.org/10.1121/1.421104>
- 770 Snowdon, C. T., & Cleveland, J. (1980). Individual recognition of contact calls by pygmy
771 marmosets. *Animal Behaviour*, *28*(3), 717–727. [https://doi.org/10.1016/S0003-](https://doi.org/10.1016/S0003-3472(80)80131-X)
772 [3472\(80\)80131-X](https://doi.org/10.1016/S0003-3472(80)80131-X)
- 773 Stepien, L. S., & Cordeau, J. P. (1960). Memory in Monkeys for Compound Stimuli. *The*
774 *American Journal of Psychology*, *73*(3), 388. <https://doi.org/10.2307/1420176>
- 775 Suied, C., Agus, T. R., Thorpe, S. J., Mesgarani, N., & Pressnitzer, D. (2014). Auditory gist:

776 Recognition of very short sounds from timbre cues. *The Journal of the Acoustical*
777 *Society of America*, 135(3), 1380–1391. <https://doi.org/10.1121/1.4863659>

778

779

780

781