The songs of the indris (Mammalia: Primates: Indridae): contextual variation in the long-distance calls of a lemur

(Article begins on next page)

23 July 2024

1   The 'songs' of the indris! Contextual variation in long distance calls in a lemur.

2   *Authors*: Valeria Torti[a], Marco Gamba[a], Zo Heritsimba Rabemananjara[b], Cristina

3   Giacoma[a]

4   [a] Department of Life Sciences and Systems Biology, University of Torino, Italy.

5   [b] Gestion de Ressources Naturelles et Environnement, University of Toamasina,

6   Madagascar.

7

8   *Corresponding author*: Valeria Torti, Department of Life Sciences and Systems Biology,

9   University of Torino, Via Accademia Albertina 13, 10123 Torino, Italy.

10   Tel. +390116704559

11   Fax. +390116704508

12   Email: valeria.torti@gmail.com

13

17

18   **ABSTRACT**

19   Contextual variation in the loud calls of strepsirhine primates is poorly understood. To

20   understand whether songs given by indris in different contexts represented acoustically

21   distinct variants and had the potential to elicit context-specific behaviours in conspecific

22   listeners, we investigated the acoustic variability of these songs and the distance

23   travelled by vocalizers after their emissions. Songs of 41 individuals were recorded

24   from 16 indri groups in 4 different forest sites in eastern Madagascar. We collected a

25   total of 270 duets and choruses arising during territorial defence, advertisement and

26   cohesion. We demonstrated that the structure of indri's songs conveyed context-specific

27   information through their overall duration, but shared the sequential pattern of harsh

28   units (roars) followed by long notes and, finally, descending phrases. Analysing in detail

29   the individual contribution to advertisement songs and cohesion songs, we found that

30   the acoustic structure of units could be classified correctly with a high degree of

31   reliability (96.23% of long notes, 80.16% of the descending phrases, 72.54% of roars).

32   Future investigations using playback stimuli could explore the relationship between

33   acoustic features and the information transmitted by the song.

34

INTRODUCTION

Loud calls are produced by many insects and vertebrates living in habitats with limited visibility and are acoustically adapted for long distance transmission (e.g. reviews by Ryan & Kime 2003; Bradbury & Vehrencamp 2011). Loud vocalizations are among the most distinctive sounds in the repertoires of non-human primates and have been the subject of numerous studies: *Cercocebus* and *Papio* (Waser 1982), *Hylobates syndactilus* (Geissmann 2000) and *Hylobates* spp. (Mitani 1992).

Different contexts are associated with the emission of loud calls: predator localization (Ghazanfar & Santos 2003), advertisement (*Hylobates* spp., Clarke et al. 2006), resource and territorial defence (Marler 1965; *Hylobates* and *Symphalangus* spp., Tembrock 1974, Mitani & Stuht 1998), and distance regulation between groups (*Alouatta palliata palliata*, Whitehead 1989) and individuals (*Indri indri*, Pollock 1986). Moynihan (1970) described how Geoffroy's tamarins (*Saguinus geoffroyi*) use long whistle vocalizations in both spacing and cohesion. Research on several other primate and non-primate species has found context-specific acoustic variation in calls [e.g. primates: Seyfarth et al. 1980; Gouzoules et al. 1984; Macedonia 1990; Rendall et al. 1999; Zuberbühler 2000; birds: Gyger et al. 1987; other mammals: Feighny et al. 2006 (elks); Theis et al. 2007 (hyenas)].

Among lemurs, indris (*Indri indri*) represent an interesting species for studying vocal communication because of their rich repertoire (Maretti et al. 2010) and the impressive long distance communication calls (known as "the song of the indri", Sorrentino et al. 2013). Indris are also the only lemurs to emit long modulated songs. The song is a complex sequence of utterances emitted by group members, males and females, adults and subadults, in a co-ordinated manner. This is a rare characteristic among primates,

59   and has evolved independently in few other genera (*Tarsius*, *Callicebus*, *Hylobates*,

60   *Nomascus* and *Symphalangus*, Geissmann 2000).

61   Several functions have been proposed for the indri's song. Pollock (1986) suggested that

62   songs serve to inform neighbouring groups about the occupation of a territory

63   (*advertisement* songs). When groups meet at territorial borders, they engage in vocal

64   contests (*territorial encounter* songs). Moreover, Pollock (1986) proposed that the song

65   has a *cohesion* function for group members, enabling the animals to be in contact when

66   they are more than 100 m apart. Additionally, the indri's song relays information about

67   group composition (sex and age of singers; Giacoma et al. 2010; Sorrentino et al. 2013),

68   and is likely to play an important role in partner localization to form new groups

69   (Pollock 1986; Giacoma et al. 2010). If these interpretations are correct, then indri loud

70   calls serve for context-related inter- and intra-group communication, with potential,

71   functionally referential implications.

72   We hypothesized that songs conveying context-specific information to conspecifics

73   should show contextual variation in their acoustic structure. To test this hypothesis, we

74   began by describing the overall pattern of songs given during the three contexts

75   described by Pollock (1986) on the basis of the Fundamental frequency (F0) profile,

76   analysing each individual animal contribution using a resynthesis-and-segmentation

77   technique (Giacoma et al. 2010). We then analysed whether advertisement songs

78   differed from cohesion songs either in their note assemblage or in their acoustic

79   structure, or both. Acoustic differences in the structure of the song would provide a

80   potential basis for modifying the behaviour of conspecific listeners. To verify whether

81   the conspecifics' behaviour changed following song emissions, we monitored individual

82   movements after advertisement and cohesion songs. We predicted that the distance

83 travelled after hearing songs emitted in different contexts would be related to

84 information encoded in the song's acoustic structure, as has been demonstrated for other

85 species (e.g. Manser et al. 2002; Spillmann et al., 2010; Fitch 2012).

86

87 **Methods**

88

89 *Study sites and subjects*

90 We observed and recorded a total of 16 groups in 4 areas of dense tropical forest in

91 Madagascar (Figure 1): 8 groups in Andasibe-Mantadia National Park, Analamazaotra

92 Reserve (18° 56' S, 48° 25' E), 3 groups in Mitsinjo Station Forestière (18° 56' S, 48°

93 24' E), 2 groups in Anjozorobe-Angavo Forest (18° 16' S, 47° 59' E) and 3 groups in

94 Maromizaha Forest (18° 56' 49'' S, 48° 27' 53'' E).

95 [FIGURE 1 HERE]

96 The indri is considered a monogamous species, usually living in groups composed of an

97 adult pair and their offspring. Study groups consisted of 2 to 8 individuals ($N_{tot} = 66$),

98 usually an adult pair, and sub-adults and juveniles of different ages. In eight groups, we

99 also recorded the presence of an additional adult individual, male or female (Table I),

100 usually, but not necessarily related with the pair. The ages of the animals studied were

101 documented by following individuals from 2004.

102

103 *Sampling and equipment*

104 We collected data in the field over a total of 21 months (3 months a year), between

105 September and December from 2004 to 2010. We carried out observations of one group

106 per day from 06.00 am to 1:00 pm. Four researchers were involved in the data

107 collection. Each researcher performed focal animal sampling on a different individual.

108 Each indri was individually recognized by natural markings. All recordings were made

109 without the use of playback stimuli and nothing was done to modify the behaviour of

110 the indris. For this research we have received permits since 2004 from "Direction des

111 Eaux et Forêts" and "Madagascar National Parks" (formerly ANGAP).

112 We recorded 270 songs, consisting of duets and choruses, with a maximum of six

113 individuals singing co-ordinately. All utterances were recorded at a distance from 2 to

114 10 m, because all the study groups were habituated. Focal animal sampling allowed the

115 attribution of each vocalization to a signaller. Recordings were made using Sennheiser

116 shotgun ME 66 and ME 67 and AKG CK 98 microphones, facing the focal animals

117 (Altmann 1974) during the emission of the songs. The microphone output signal was

118 recorded using a solid-state digital audio recorder Marantz Portable (PMD671) at a

119 sampling rate of 44.1 kHz.

120

121 *Contexts of emission*

122 Following Pollock (1986), we identified and categorised three behavioural contexts

123 during which indri diurnal singing took place. We labelled songs emitted during

124 intergroup encounters as "territorial encounter songs" (hereinafter TSs, Figure 2), when

125 there was visual contact between individuals of two different social groups. We labelled

126 songs emitted while animals were feeding or resting in the trees, or interacting socially,

127 as "advertisement songs" (hereinafter ASs, Figure 2). Songs given when individuals

128 from the same group were not in visual contact were labelled as "cohesion songs"

129 (hereinafter CSs, Figure 2). When following indri groups in the field, we labelled each

130 song according to the position of the focal group in relation to neighbouring groups and

131 the spatial relationships between group members (Table I).

132 [TABLE I HERE] [FIGURE 2 HERE]

133 *Definition of the vocal units*

134 We used the term "song" for the complex of roar followed by a sequence of modulated

135 notes (Sorrentino et al. 2013). We used the term "call" to identify the individual

136 contribution to the song. We indicated as "notes" or "units" the single sounds

137 constituting the modulated part of the song (Thalmann et al. 1993), distinguishing

138 between "long notes" emitted in the first part and "DP notes" for the units comprising a

139 descending phrase. "Descending phrases" (Thalmann et al. 1993) are sequences of notes

140 in which the start of a successive note is characterized by a F0 lower than the final F0 of

141 the previous one. Notes within a descending phrase are emitted at less than 0.8 s from

142 one another (Gamba et al. pers. obs.). We termed songs given by two singers "duets",

143 while the term "chorus" referred to more than two indris contributing to the song.

144

145 *Acoustic analyses*

146 Songs emitted in the three contexts consisted of roars followed by a long sequence of

147 harmonic notes. In this harmonic sequence long notes occurred first and were followed

148 by the descending phrases (Figure 3; Thalmann et al. 1993; Giacoma et al. 2010).

149 During the song the simultaneous singing of all individuals produced an overlapping of

150 notes. The extensive overlapping of notes during TSs (Figure 2) did not allow the

151 extraction of group and individual contributions to these songs; hence, we only

152 measured song duration in five TSs. We were, however, able to extract individual

153 contributions and, consequently, to carry out detailed quantitative analyses of note

154 occurrences and of temporal and frequency parameters in CSs and ASs.

155 We edited segments containing indri songs using Praat 5.2.26 (Boersma & Weenink

156 2008) and copied each song to a single audio file (in AIFF format). A silent period of

157 0.5 s was inserted at the beginning and the end of each sound file. Vocal signals were

158 named after song vocalizers (Sorrentino et al. 2013). We then divided each individual

159 emission according to previously set labels and text grids into single files using custom-

160 made software (Giacoma et al. 2010). Using a resynthesis-and-segmentation technique

161 we focused on the study of the F0 variation over time, which allowed us, by means of

162 the autocorrelation pitch contour method, to separate out the contribution of each

163 individual participating in the song. Using a pitch extraction algorithm, Praat generates

164 F0 contours that can be treated as a linear sequence of points, defined in the temporal

165 and frequency domains. The F0 contour of each note emitted by the focal animal was

166 extracted in steps of 0.01 s ("to pitch cc": time step 0.01; silence threshold 0.10;

167 minimum F0 150 Hz; maximum F0 1 800 Hz); the F0 contours were visually inspected

168 to avoid any incorrect candidates entering the contour, and saved to a separate file. On

169 the basis of the F0 contour of the note sequence of the focal individual we then

170 resynthesized the utterance using the "To Sound (sine)" function in Praat. We obtained a

171 new sound featuring one emitter's notes and then segmented this utterance into files

172 containing single notes. For each note, we extracted the following parameters of the F0

173 contour: duration, average (MeanF0), maximum (MaxF0) and minimum (MinF0),

174 StartF0, EndF0, EnstF0, Pt2Max, Pt2Min (Table IIa). The final output file was

175 assembled within Praat and exported to a Microsoft© Excel spreadsheet (Gamba &

176   Giacoma 2007; Gamba et al. 2011). F0 parameters were calculated for each unit using

177   typical settings: F0 range 150 – 1800 Hz; autocorrelation method - automatic;

178   maximum number of candidates 35. The individual contribution to the song was also

179   described in relation to the overall song features (Figure 3). We measured the structural

180   and temporal features reported in Table IIb.

181   [TABLE II HERE] [FIGURE 3 HERE]

182

183   *Statistical Analyses*

184   We quantitatively analysed the acoustic structure of 235 advertisement songs and 21

185   cohesion songs for a total of 41 singing individuals, adults and sub-adults, 21 males and

186   20 females. We reported means and standard deviations (± SD) of the acoustic variables

187   measured. We used the Mann-Whitney test to analyse context-related differences in the

188   overall pattern of the song because the data were not normally distributed, and

189   Wilcoxon tests to analyse potential differences in the number of roars, long notes and

190   descending phrase notes emitted by the same individual in the two different contexts.

191   Using acoustic parameters averaged across songs, we applied a linear model (GLM) to

192   test whether the context of emission and sex of the caller had an effect on song

193   parameters "Calldur" and "Percdur" (using *glm* in R, R Development Core Team 2008).

194   We conducted stepwise Discriminant Function Analysis (sDFA) to test whether note

195   types uttered in ASs and CSs could be correctly assigned to the context in which the

196   respective songs were emitted. We also identified which parameters were most

197   important to discriminate ASs from CSs. To avoid pseudoreplication and to validate the

198   classificatory analyses on a congruent number of advertisement and cohesion songs, we

199   randomly selected subsets of ASs and CSs. Random subsets were obtained using a

200   custom syntax in IBM SPSS Statistics 19 for Mac (IBM SPSS Inc., USA). Results

201   reported for the sDFA are average values of those obtained from the each subset run.

202   Ranges of the respective P values are also given. The percentage of correctly classified

203   instances of a cross-validated (c.-v.) discriminant model is reported for each analysis.

204   All the above-mentioned statistical analyses were performed using IBM SPSS Statistics

205   19 for Mac.

206   We used generalized linear mixed models (GLMM) to analyse the effect of sex and

207   context on single call parameters. We used this model because it accounts for unequal

208   sample sizes and for repeated observations of the same subjects by including the subject

209   (using an individual identity code) as a random factor in the model (Pinheiro & Bates

210   1996). Context of emission (Advertisement = 0, Cohesion = 1) and sex of the caller

211   (Male = 1, Female = 2) were entered in the model as fixed factors. We tested the data of

212   call parameters for normal distribution using the Shapiro-Wilk test. Because the data

213   were not normally distributed they were log-transformed before analysis. GLMM were

214   performed using *glm* and *nlme* in R (R Development Core Team 2008).

215

216   *Behavioural responses*

217   To describe responses to ASs and CSs quantitatively, we noted whether indris moved

218   after the emission of the song and measured the distance travelled along a continuous

219   path by the focal animal within 10 minutes of the call. We estimated the horizontal

220   distance by visual tracking within 5 m or calculated the distance travelled by the focal

221   animal using GPS waypoints as a straight-line distance. To verify the hypothesis that

222   CSs promoted reunification of the group, we hypothesized that after CSs individuals

223   would move to join the distant members of the group, and that therefore the average

224   distance travelled by individuals after the emission of CSs should be significantly

225   longer than after the emission of ASs. We applied the Mann-Whitney U test because the

226   data were not normally distributed.

227

228   **RESULTS**

229

230   *Context, frequency of emission and total duration of the songs*

231   We observed the emission of TSs during active territorial defence, when social groups

232   came into direct contact at a shared boundary, and indris initiated a direct confrontation

233   by singing towards the opponents. When emitting TSs, the indris faced one another with

234   the head orientated towards the neighbours. All group members participated in these

235   long choruses, singing was very hectic and notes emitted by the different individuals

236   overlapped extensively (Figure 2). We recorded 14 TSs (or 0.05 TSs per day), but only

237   in 5 were we able to recognize singing of the focal group. After intense vocal fights

238   these encounters could either lead or not lead to physical aggression. Contacts were

239   terminated by the two groups disappearing rapidly in opposite directions.

240   The ASs were recorded while animals were feeding or resting in the trees, or interacting

241   socially. They could start singing spontaneously or in response to songs emitted by other

242   indri groups. During these songs all individuals within a social group were in visual

243   contact and singers frequently turned their heads in different directions. ASs were the

244   most frequently recorded songs (2.22 per day).

245 The CSs were emitted when individuals were spending time away from their social

246 groups. In these situations, indris often uttered songs while orientating their heads in the

247 direction in which other group members were heard singing. CSs were characterized by

248 scant overlap between notes (Figure 2), and were recorded at a mean rate of 0.07 per

249 day.

250 The average duration of TSs (685.32 ± SD 338.97 s, N = 5) was markedly longer than

251 those of the other songs. They were nearly five times longer than the AS average (N =

252 235, 112.83 s ± SD 53.46 s) and ten times longer than the CS average (N = 21, 70.68 s ±

253 SD 20.18 s). ASs were significantly longer than CSs (Mann-Whitney: U = 1022.500, W

254 = 1253.500, Z = -4.402, p < 0.001).

255

256 *Context- and sex-related differences in individual contributions to the song*

257 The GLM analyses showed that the factor "context" had a significant effect on the total

258 duration of individual contributions to the song (F = 11.5666; P < 0.001), while sex had

259 a significant effect on the percentage of individual contributions relative to song

260 duration (F = 16.6300; P < 0.001). The duration of individual contributions to the song

261 reflected the result found for song duration, with males singing longer (77.48 s ± SD

262 9.95 s) when emitting ASs as opposed to CSs (59.77 s ± SD 14.43 s, df = 7, t = 2.661, P

263 = 0.032). The ratio of song duration represented by individual males calling also varied

264 significantly in the two singing contexts: from 74.18 s ± SD 18.45 s in ASs to 89.07 s ±

265 SD 20.06 s in CSs (t = -2.585, df = 7, P = 0.036). Female indris sang an average of

266 70.17 s ± SD 32.32 s when emitting ASs as opposed to CSs (64.18 s ± SD 18.56 s, t =

267 0.632, df = 6, P = 0.550). Similarly, individual female ratios did not differ with the

268 context: 81.49 s ± SD 30.85 s in ASs; 89.48 s ± SD 16.16 s in CSs (t = -1.373, df = 6, P

269 = 0.219). ASs contained a significantly higher number of notes (24.69 ± SD 9.03) than

270 were contained in CSs (18.43 ± SD 8.92, t-test: t = 2.232, P = 0.031). When sexes were

271 tested separately we found that only males showed a higher number of notes in ASs

272 (ASs: 20.72 ± SD 5.36, CSs: 14.75 ± SD 7.05, t-test: t = 2.209, P = 0.040, N = 10).

273 When considering inter-note silences, we observed that ASs showed on average 54.89 s

274 ± SD 28.12 s of unvoiced intervals against 30.58 s ± SD 14.82 s in CSs (t-test: t =

275 3.319, P = 0.002). The longer duration of individual contributions in ASs was based on

276 significantly longer cumulative inter-note silences, which emerged clearly when sexes

277 were tested separately.

278

279 *Context-related differences in song composition*

280 To investigate whether songs emitted in advertisement and cohesion contexts differed in

281 structure, we compared the likelihood of transition from one phase to the other: ASs

282 started with roars in 84.25% of cases against 66.67% in CSs; descending phrases were

283 introduced by long notes in 98.72% of cases in ASs, and in 47.62% of cases in CSs.

284 ASs and CSs differed in note assemblage (Table III) at the individual level (in order to

285 account for sex differences). Roars were significantly more frequent in male

286 contributions to ASs than to CSs (Table III; Wilcoxon test: Z = -2.023, P = 0.043, N=5),

287 but this was not true of the contributions of females (Wilcoxon test: Z = -0.405, P =

288 0.686, N = 6). The number of long notes (Table III; Wilcoxon test for males: Z = 0.943,

289 P = 0.345, N = 7; this analysis could not be performed for females because long notes

290 are less frequent in female contributions) and descending phrase notes (Table III;

291 Wilcoxon test for males, Z = -1.260, P = 0.208, N = 8; Wilcoxon test for females: Z = -

292 0.338, P = 0.735, N = 7) was not significantly different in ASs and CSs.

293   The GLMM analyses showed that sex affected Pt2Min and Pt2Max for roars, long notes

294   and descending phrase units (Table III). Context affected MaxF0 and StartF0 (Table

295   IV).

296   [TABLE III HERE] [TABLE IV HERE]

297

298   *Context-related differences in the acoustical structure of notes*

299   We tested the hypothesis that songs given in different contexts could comprise vocal

300   units with different acoustical structures. Stepwise DFA with randomly chosen subsets

301   showed that the most important variables involved in discriminating roars were EnstF0

302   and MinF0. The model classified 72.54% ± 5.99% (c.-v. 69.98% ± 5.46%, Wilk's

303   Lambda = 0.72 ± 0.11, 0.001< P ≤ 0.05) of the roars correctly. Models discriminating

304   long notes always involved either duration, EnstF0 or Pt2.Max, and classified 96.23% ±

305   3.48% (c.-v. 95.15% ± 4.05%, Wilk's Lambda = 0.23 ± 0.05, P < 0.001) of the cases

306   correctly. Models discriminating modulated notes in descending phrases relied mostly

307   on minF0 and duration, and correctly classified 80.16% ± 8.22% (c.-v. 78.81% ± 7.60%,

308   Wilk's Lambda = 0.59 ± 0.09, 0.001 < P < 0.011) of the cases.

309

310   *Locomotor behaviour following song emission*

311   To understand whether contextual differences in song structure have significance for

312   indris, we observed their reactions to the songs uttered, and quantified their movements

313   following ASs and CSs in terms of distance travelled by group members. Only in 46.1%

314   of cases did indris move after ASs. When moving, individuals usually ascended or

315   descended while remaining in the same tree, or jumped on to an adjacent one. After ASs

316 they covered on average 7.41 m ± SD 20.38 m (range = 0-100 m). Indris emitted CSs

317 directing their calls at the distant group member(s). They were always out of sight of

318 one another and separated by over 50 m. Indris always left the calling tree after emitting

319 CSs. They travelled on average 55.82 m ± SD 46.56 m away from the tree, usually

320 moving horizontally towards the other callers and re-uniting with group members that

321 had been out of visual contact. The distance travelled after CSs (range = 7-173 m) was

322 significantly longer than that travelled after ASs (Mann-Whitney: U = 42.000, P <

323 0.001, N = 55). In the case of both ASs (82.9 %; N_{males} = 14) and CSs (70.4 %; N_{males} =

324 6) males started to call first and females soon overlapped them. Sub-adults of both sexes

325 started singing CSs in 14.8 % of cases (3 females and 1 male), while this percentage

326 was 22.2 % for ASs.

327

328 **DISCUSSION**

329 We have provided strong evidence that indri songs are not acoustically homogeneous

330 signals, but show changes in duration, note assemblage and acoustical structure that

331 elicit distinct responses from group members. This is consistent with what has been

332 observed in gibbons (*Symphalangus syndactylus* and *Hylobates* spp.) that showed

333 distinct song types. In gibbons, which share similarities with indris in terms of both

334 territorial behaviour and social structure, calls usually exchanged across territorial

335 boundaries serve in maintaining spacing between neighbours (Chivers 1972, 1974,

336 1976; Cowlishaw 1992), but when delivered within the range of another conspecific

337 pair, these signals are followed by distance reduction and direct chasing between males

338 (Mitani 1985; Reichard & Sommer 1997). Our results support the spacing function

339 hypothesis in many respects: (i) just before starting the emission of a song, indris moved

340 up the trees to where vegetation was less dense and attenuation was consequently lower;

341 (ii) they sang with their mouths orientated towards the territories of all the neighbouring

342 groups (Favaro et al. 2008; Gamba et al. 2011); and (iii) after the emission of ASs indris

343 generally moved up and down on the branches or in the surrounding trees. This

344 behaviour is consistent with the territory defence hypothesis (Ellefson 1974; Tenaza

345 1985).

346 We have demonstrated quantitatively that CSs and ASs differ in duration because of a

347 significant increase in the silent inter-note intervals in ASs, and in the duration of the

348 male contributions. A longer duration may influence the way a signal maximizes its

349 detectability (Guilford and Dawkins 1991). Thus longer units and a song of greater

350 length can maximize the probability of detection by conspecifics that are outside the

351 territorial boundaries (Pollock 1986) and that are only sporadically attentive (Wiley

352 1983). In fact, in the case of ASs, neighbours are usually busy carrying out their normal

353 daily activities. The longer duration of ASs versus CSs may also reflect a higher

354 investment in communicating group's presence in the territory to avoid intrusions. This

355 supports Pollock's (1979) hypothesis that ASs may have a warning function.

356 We found that males units are significantly higher in frequency in ASs than in CSs, and

357 are longer than those of females in ASs. These makes sense in the light of the male's

358 predominant role in defending the territory (Pollock 1979). This structural adaptation

359 also found support in the fact that roars, and especially male roars, are more frequently

360 given in ASs. The roars' prominent chaotic pattern may explain their being relatively

361 invariant across contexts when compared to the other unit types. As suggested by

362 previous studies (Pollock 1986, Maretti et al. 2010) they really are distinct from the rest

363 of the song.

364 ASs showed a higher degree of stereotypy in the sequence of notes, which may also

365 increase song detectability (Rogers & Kaplan 2000). Changes in units' assemblage and

366 occurrence parallel the evidence from studies of Sumatran orangutans, where contextual

367 differences are expressed through temporal and structural variation of the signals rather

368 than by the types of units uttered (Spillmann et al. 2009). Such combinatorial signalling

369 has also been described in birds and other primates with relatively small vocal

370 repertoires (Zuberbühler 2003; Catchpole 2010).

371 Our data on the emission of CSs confirms Pollock's (1986) hypothesis that singing may

372 also gather group members together. These findings also complement earlier research on

373 primate species (Deputte 1973, 1978; De Vore & Hall 1965; Waser 1977) showing that

374 finely-tuned communicative abilities, such as those required to produce a complex song,

375 may also serve a intra-group cohesion function.

376 In our study we confirmed that acoustical cues to sex are encoded in the indris song in

377 both temporal and frequency parameters (Giacoma et al. 2010). It is likely that the

378 indris' song has the potential to provide conspecifics with information about the

379 emitter's sex and the composition of the vocalizing group (e.g. number of adult males vs

380 females; Mitani and Stuht 1998; Mendez-Cardenas et al. 2008).

381 Following Marshall and Marshall (1976) we suggest that different selection pressures

382 acted on male and female contributions also in indris. In indris, males singing in the ASs

383 showed both a higher frequency in long notes and a lower frequency in descending

384 phrases. Different parts of the same individual's contribution may differ in function

385 (Goustard 1985), as has been demonstrated for birds duets (Sonnenschein & Reyer

386   1983).

387

388   **Acknowledgements**

410 manuscript. The contents of this document are the sole responsibility of the authors and

411 can under no circumstances be regarded as reflecting the position of the European

412 Union.

413

414 **References**

415 Altmann J. 1974. Observational study of behavior: sampling methods. Behaviour

416 49:227-267.

417 Boersma P, Weenink, D. 2008. Praat: doing phonetics by computer (version 5.0 27).

418 Computer program. University of Amsterdam. The Netherlands.

419 Bradbury JW & Vehrencamp SL. 2011. Principles of Animal Communication (2nd

420 edition). Sunderland MA: Sinauer Associates.

421 Catchpole C. 2010. The functions of advertising song in the sedge warbler

422 (*Acrocephalus schoenobaenus*) and the reed warbler (*A. scirpaceus*). Behaviour 46:

423 300-320.

424 Chivers DJ. 1972. The siamang and the gibbon of the Malay peninsula. Gibbon and

425 Siamang 1:103-135.

426 Chivers DJ. 1974. The siamang in Malaya: a field study of a primate in tropical rain

427 forest. Contributions to Primatology 4:1-335.

428 Chivers DJ. 1976. Communication within and between family groups of Siamang

429 (*Symphalangus syndactylus*). Behaviour 57:116-135.

430 Clarke E, Reichard UH, Zuberbühler K. 2006. The syntax and meaning of wild gibbon

431 songs. PLoS ONE 1, 1:e73.

432 Cowlishawn G. 1992. Song function in gibbons. Behaviour 121:131-153.

433 Deputte BL. 1973. Étude d'un type de comportement vocal chez un groupe captif de

434 mangabeys (*Cercocebus albigena*, Gray). Méthode télémetrique d'enregistrement

435 individuel des vocalisations. DEA Biologie Animale: Eco-Ethologie. Rennes: Université

436 de Rennes I.

437 Deputte BL. 1978. Étude du répertoire vocal du Gibbon a favoris blancs (*Hylobates*

438 *concolor leucogenys*): Analyse structurale des vocalisations. Zeitschrift für

439 Tierpsychologie 48:225-250.

440 De Vore I, Hall KRL. 1965. Baboon ecology. In: De Vore I, editor. New York: Holt

441 Reinhart and Winston.

442 Ellefson JO. 1974. A natural history of white-handed gibbons in the Malayan Peninsula.

443 Gibbon and Siamang 3:1-136.

444 Favaro L, Gamba M, Sorrentino V, Torti V, Giacoma C. 2008. Singers in the Forest:

445 Acoustic Structure of indri's loud calls and first evidence of vocal tract tuning in a

446 prosimian primate. In: Proceedings of the XXXV Annual Congress of the Italian

447 Acoustic Association (AIA). Milano: Associazione Italiana di Acustica.

448 Feighny J, Williamson K, Clarke J. 2006. North American elk bugle vocalizations: male

449 and female bugle call structure and context. Journal of Mammalogy 87:1072–1077.

450 Fitch WT. 2012. Segmental structure in banded mongoose calls. BMC Biology 10:98.

451   Gamba M, Giacoma C. 2007. Quantitative acoustic analysis of the vocal repertoire of
452   the crowned lemur. Ethology Ecology And Evolution 19:323-343.

453   Gamba M, Favaro L, Torti V, Sorrentino V, Giacoma C. 2011. Vocal tract flexibility and
454   variation in the vocal output in wild indris. Bioacoustics 20:251-266.

455   Geissmann T. 2000. Duet songs of the Siamang *Hylobates syndactilus*: I structure and
456   organization. Primate Reports 56:33-60.

457   Ghazanfar AA, Santos LR. 2003. Primates as auditory specialists. In: Ghazanfar AA,
458   editor. CRC Press. pp. 1-12.

459   Giacoma C, Sorrentino V, Rabarivola C, Gamba M. 2010. Sex differences in the song of
460   *Indri indri*. International Journal of Primatology 31:539-551.

461   Goustard M. 1985. Structure acoustique et fonctions des vocalisations territoriales chez
462   le Gibbon à mains blanches (*Hylobates lar*) observé dans son habitat naturel en
463   Thaïlande. Annales des Sciences Naturelles Zoologie 13:265-279.

464   Gouzoules S, Gouzoules H, Marler P. 1984. Rhesus monkey (*Macaca mulatta*) screams:
465   representational signalling in the recruitment of agonistic aid. Animal Behaviour
466   32:182–193.

467   Guilford T, Dawkins, MS. 1991. Receiver psychology and the evolution of animal
468   signals. Animal Behaviour 42:1-14.

469   Gyger M, Marler P, Pickert R. 1987. Semantics of an avian alarm call system: the male
470   domestic fowl, *Gallus domesticus*. Behaviour 102:15–40.

471   Macedonia JM. 1990. What is communicated in the antipredator calls of lemurs:

472   Evidence from playback experiments with ringtailed and ruffed lemurs. Ethology

473   86:177-190.

474   Manser MB, Seyfarth RM, Cheney DL. 2002. Suricate alarm calls signal predator class

475   and urgency. Trends in Cognitive Sciences 6:55-57.

476   Maretti G, Sorrentino V, Finomana A, Gamba M, Giacoma C. 2010. Not just a pretty

477   song: an overview of the vocal repertoire of Indri indri. Journal of Anthropological

478   Sciences 88:151-165.

479   Marler P. 1965. Communication in monkeys and apes. In: De Vore I, editor. New York:

480   Holt Rinehart, Winston. pp 420-438.

481   Marshall JT, Marshall ER. 1976. Gibbons and their territorial songs. Science 193:235-

482   237.

483   Mendez-Cardenas M, Randriananbinina B, Rabesandratana A, Rasoloharijaona S,

484   Zimmermann E. 2008. Geographic variation in loud calls of sportive lemurs (Lepilemur

485   spp.) and their implications for conservation. American Journal of Primatology 70: 1–

486   11.

487   McComb K, Semple S. 2005. Coevolution of vocal communication and sociality in

488   primates. Biology Letters 1:381-385.

489   Mitani JC. 1985. Gibbon song duets and intergroup spacing. Behaviour 92:59-96.

490   Mitani JC. 1992. Singing behavior of male gibbons: field observations and experiments

491   In: Nishida T, McGrew W, Marler P, Pickford M, DeWaal F. Tokyo: University of Tokyo

492   Press. pp. 199-210.

493   Mitani JC, Stuht J. 1998. The evolution of nonhuman primate loud calls: acoustic

494   adaptation for long-distance transmission. Primates 39:171-182.

495   Moynihan M. 1970. Some behavior patterns of playrrhine monkeys II. *Saguinus*

496   *geoffroyi* and some other tamarins. Smithsonian Contributions to Zoology 28:1–76.

497   Pereira ME, Seeligson ML, Macedonia JM. 1988. The behavioral repertoire of the

498   black-and-white ruffed lemur, Varecia variegata variegata (Primates: Lemuridae). Folia

499   Primatologica 51:1-32.

500   Petter JJ, Albignac R, Rumpler Y. 1977. Mammifères lémuriens. Faune de Madagascar

501   44:391–410.

502   Petter JJ, Charles-Dominique P. 1979. Vocal communication in prosimians. In: Doyle

503   GA, Martin RD, editors. New York. Academic Press. pp. 272–282.

504   Pinheiro JC, Bates DM. 1996. Unconstrained parameterizations for variance-covariance

505   matrices. Statistics and Computing 6:289–296.

506   Pollock JI. 1979. Spatial distribution and ranging behaviour in lemurs. In: Doyle GA,

507   Martin RD, editors. New York: Academic Press. pp. 359-409.

508   Pollock JI. 1986. The song of the indris (Indri indri; Primates: Lemuroidea): natural

509   history form and function. International Journal of Primatology 7:225-264.

510 Reichard U, Sommer V. 1997. Group encounters in wild gibbons (*Hylobates lar*):

511 agonism affiliation and the concept of infanticide. Behaviour 134:1135-1174.

512 Rendall D, Cheney DL, Seyfarth RM, Owren MJ. 1999. The meaning and function of

513 grunt variants in baboons. Animal Behaviour 57:583-592.

514 Rogers LJ, Kaplan G. 2000. Songs, Roars, and Rituals: Communication in Birds,

515 Mammals, and Other Animals. Harvard University Press, Cambridge, MA. pp. 1-207.

516 Ryan MJ, Kime NM. 2003. Selection on long distance acoustic signals In: Simmons A,

517 Fay R, Popper A, editors. Berlin: Springer Verlag. pp. 225-274.

518 Seyfarth RM, Cheney DL, Marler, P. 1980. Monkey responses to three different alarm

519 calls: evidence of predator classification and semantic communication. Science

520 210:801-803.

521 Sonnenschein E, Reyer HU. 1983. Mate-guarding and other functions of antiphonal

522 duets in the slate-coloured boubou (*Laniarius funebris*). Zeitschrift für Tierzuchtung

523 und Zuchtungsbiologie 63:112-140.

524 Sorrentino V, Gamba M, Giacoma C. 2013. A quantitative description of the vocal types

525 emitted in the indri's song. In: Masters JC, Gamba M, Genin F, editors. Leaping Ahead:

526 Advances in Prosimian Biology. New York: Springer. p. 315-322.

527 Spillmann B, Dunkel LP, Van Noordwijk MA, Amda RNA, Lameira AR, Wich SA, Van

528 Schalk CP. 2010. Acoustic properties of long calls given by flanged male Orang-Utans

529 (*Pongo pygmaeus wurmbii*) reflect both individual identity and context. Ethology

530 116:385-395.

531   Tembrock G. 1974. Sound production in Hylobates and Symphalangus In: Rumbaug

532   DM, editor. Gibbon and Siamang. Basel: Karger. pp. 176-205.

533   Tenaza R. 1985. Songs of Hybrid-Gibbons (*Hylobates lar* x *H. muelleri*). American

534   Journal of Primatology 8:249-253.

535   Thalmann U, Geissmann T, Simona A, Mutschler T. 1993. The indris of Anjanaharibe-

536   Sud Northeastern Madagascar. International Journal of Primatology 14:357-381.

537   Theis KR, Greene KM, Benson-Amram SR, Holekamp KE. 2007. Sources of variation

538   in the long-distance vocalizations of spotted hyenas. Behaviour 144:557-584.

539   Waser PM. 1977. Individual recognition intragroup cohesion and intergroup spacing:

540   evidence from sound playback to forest monkeys. Behaviour 60:28–74.

541   Waser PM. 1982. The evolution of male loud calls among mangabeys and baboons In:

542   Brown CH, Petersen MR, Snowdon CT, editors. New York: Cambridge University

543   Press. pp. 117-143.

544   Whitehead JM. 1989. The effect of the location of a simulated intruder on responses to

545   long-distance vocalizations of mantled howling monkeys (*Alouatta palliata palliata*).

546   Behaviour 108:73–103.

547   Wiley RH. 1983. The evolution of communication: information and manipulation. In:

548   Halliday TR, Slater PJB, editors. Oxford: Blackwell. pp. 156-189.

549   Zuberbuhler, K. 2000. Interspecies semantic communication in two forest primates.

550   Proceedings of the Royal Society of London Series B: Biological Sciences. 267:713-

551   718.

552 Zuberbühler K. 2003. Referential signalling in non-human primates: cognitive

553 precursors and limitations for the evolution of language. Advances in the Study of

554 Behavior 33:265-307.

Figure 1. Map of the study area showing the location of the different forest sites.

Figure 2. Representations of a territorial encounter song (TS), an advertisement song (AS) and a cohesion song (CS) as original spectrograms. Spectrogram settings: View range = 0 to 3000 Hz; Window length = 0.05 s; Dynamic range = 70 dB.

Figure 3. Graphical representation of a cohesion song as original spectrogram and as resynthesized individual contributions. We represent here the resynthesized contribution of the adult male (above) and the adult female (below). For each resynthesized call we labelled the different note types, classified as roars, long notes (LN) and descending phrases (from DP1 to DP4 in male's contribution, from DP1 to DP9 in female's contribution). Each DP unit is labelled with a number, from 1 to 6.

Tab. I. Overview of the groups and songs recorded.

group and forest	N_{tot}	composition	singers	N callers	AS	CS	TS
1 Analamazaotra	4	1AM;1AF;1SM;1J	1AM; 1AF; 1SM	2.59 ± SD 0.65	22	2	
2 Analamazaotra	8	3AM;2AF*;1SM;1SF;1J	2AM;1AF;1SF		21	1	
3 Analamazaotra	6	3AM;1AF;1SM;1J	3AM;1AF		41	1	2 with @
5 Analamazaotra	5	1AM;2AF*;2J*	1AM;2AF		17	1	
6 Analamazaotra	4	2AM; 1AF; 1AF*	1AM; 1AF		8	-	
7 Analamazaotra	2	1AM; 1AF	1AM; 1AF		11	-	
1 Mantadia	5	2 AM; 2 AF; 1J*	2 AM; 1AF		4	-	
4 Mantadia	2	1AM; 1AF	1AM; 1AF		5	-	
1 Station Forestière	6	3AM*;1AF;1SM;1J	2AM;1AF	2.58 ± SD 0.70	41	4	3 (2 with @ and one with 2 SF)
2 Station Forestière	5	1AM;3AF;1J	1AM; 2AF		40	-	7 with @
3 Station Forestière	3	1AM;1AF;1J	1AM;1AF		8	7	
1 Anjozorobe-Angavo	4	2AM*;1AF;1SF	1AM;2AF	2.18 ± SD 0.38	3	2	
2 Anjozorobe-Angavo	2	1AM;1AF	1AM;1AF		3	2	
2 Maromizaha	3	1AM;1AF; 1J*	1AM;1AF	1.95 ± SD 0.21	5	-	2 with 3MZ
3 Maromizaha	4	1AM; 1AF; 1SM; 1J	1AM;1AF		6	-	
4 Maromizaha	3	1AM;1AF;1J	1AM;		-	1	
TOT	66	66	41	-	235	21	14

Age classes according to Pollock 1986: AM = adult male (>6 years); AF = adult female (<6 years); SM = sub-adult male (3-6 years); SF = sub-adult female (3-6 years); J = juvenile (not possible to define sex); * = departed (dead animals or individuals leaving the group). "N_{tot}" represents the total number of individuals in a group, including animals not participating to the song. For each group we reported the site in which the group lives, the number of individuals (N_{tot}) and the group composition, the number and the sex of the singers, the number of advertisement songs (AS), of cohesion songs (CS) and of territorial encounter songs (TS) analysed in this study. "N callers" represents the mean number of calling individuals in each forest area. @ indicates the encounter with a non-habituated group that occurred at the territorial boundaries. We recorded one CS of the group 4MZ by chance and we did not record ASs because 4MZ was not a focus group.

Tab. II. Definitions and abbreviations of the acoustic and temporal parameters and nomenclature used in the study.

Label	Description of Parameters
a. Spectral features of notes	
Duration	Total duration of the note (s)
MeanF0	Mean value of F0 for each note (Hz)
MaxF0	Maximum value of F0 for each note (Hz)
MinF0	Minimum value of F0 for each note (Hz)
EnstF0	F0 end – F0 start (Hz)
StartF0	Value of F0 at the beginning of the note (Hz)
EndF0	Value of F0 at the end of the note (Hz)
Pt2max	% of signal duration before MaxF0
Pt2Min	% of signal duration before MinF0
b. Features of signals	
Songdur	Total (voiced + unvoiced) duration of the song (s)
Calldur	Total duration of each individual contribution (s)
Percdur	Calldur/Songdur (expressed in %)
Sil_Sum_Dur	Total duration of the silent intervals between notes (s)
N_notes	Total number of notes for each individual contribution
c. Vocal units nomenclature	
Song	The complex of roars followed by a sequence of modulated notes (Sorrentino et al. 2013). We termed songs given by two singers "duets", while the term "chorus" referred to more than two indris contributing to the song.
Call	The individual contribution to the song.
Notes/Units	The single sounds constituting the modulated part of the song (Thalmann et al. 1993).
Long Notes	The longest, modulated notes emitted in the first part of the song, immediately after the roars (or eventually following an intermediate short note).

DP Notes	The units comprised in a descending phrase. Notes within a descending phrase are emitted at less than 0.8 s from one another (Gamba et al. pers. obs.).
Descending Phrase	Sequences of notes in which the start of a successive note is characterized by a F0 lower than the final F0 of the previous one (Thalmann et al. 1993).

Tab. III. Mean values and standard deviations of each acoustic parameter in relation to sex, note type and context of emission of the songs.

Note type	Parameter	Males				Females			
		Advertisement		Cohesion		Advertisement		Cohesion	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Roars	Duration (s)	0.905	0.283	1.012	0.522	0.846	0.135	1.050	0.291
	MeanF0 (Hz)	282	32	245	45	26	30	243	28
	MaxF0 (Hz)	327	45	262	49	303	60	271	30
	MinF0 (Hz)	229	32	216	44	228	26	209	31
	StartF0 (Hz)	280*	29	231*	41	241	32	218	30
	EndF0 (Hz)	278	37	245	51	286	48	260	37
	Pt2Min (%)	26.70*	8.24	28.81*	28.67	20.60	8.64	21.49	11.90
	Pt2Max (%)	61.37	9.66	60.65	14.27	78.89*	9.58	68.41*	22.75
	EnstF0 (Hz)	69	54	16	34	55	32	44	26
	N_notes	3.1	0.8	2.3	0.4	2.4	1.0	2.2	1.0
Long notes	Duration (s)	2.788*	0.422	2.406*	1.041	1.447	0.234	1.049	0.212
	MeanF0 (Hz)	735	35	889	231	779	45	913	195
	MaxF0 (Hz)	844	45	938	219	874	61	1232	12
	MinF0 (Hz)	651	43	842*	245	736*	37	812	286
	StartF0 (Hz)	788	53	912	226	774	50	873	201
	EndF0 (Hz)	750	39	892	223	865*	61	1229*	15
	Pt2Min (%)	51.42*	11.48	54.41*	22.60	26.58	12.44	7.51	1.18
	Pt2Max (%)	35.56	10.52	38.14	18.62	84.96*	10.96	88.92*	0.25
	EnstF0 (Hz)	82.37	41.25	25.31	60.81	100.83	32.51	356.32	185.87
	N_notes	2.36	0.72	2.36	1.25	3.43	1.55	1.0	0.0
Descending phrase units	Duration (s)	1.417*	0.131	1.684*	0.248	1.135	0.130	1.174	0.127
	MeanF0 (Hz)	925*	68	856	65	886	32	858*	19
	MaxF0 (Hz)	982	53	926	60	980	39	943	33
	MinF0 (Hz)	866	39	801	68	834	34	805	19
	StartF0 (Hz)	934	47	884	68	901	31	883	32
	EndF0 (Hz)	920	46	856	59	932	46	899	30
	Pt2Min (%)	58.25*	5.55	57.95*	9.58	42.64	6.34	45.60	3.25
	Pt2Max (%)	37.64	6.13	33.95	11.16	52.68*	13.51	50.57*	6.14
	EnstF0 (Hz)	62	41	42	48	96*	36	60*	49
	N_notes	15.60	4.90	13.09	5.16	23.36*	7.52	24.69*	5.53

Asterisks denote those parameters showing the highest significant values in the GLMM across-sexes ($p<0.05$), for both advertisement and cohesion songs.
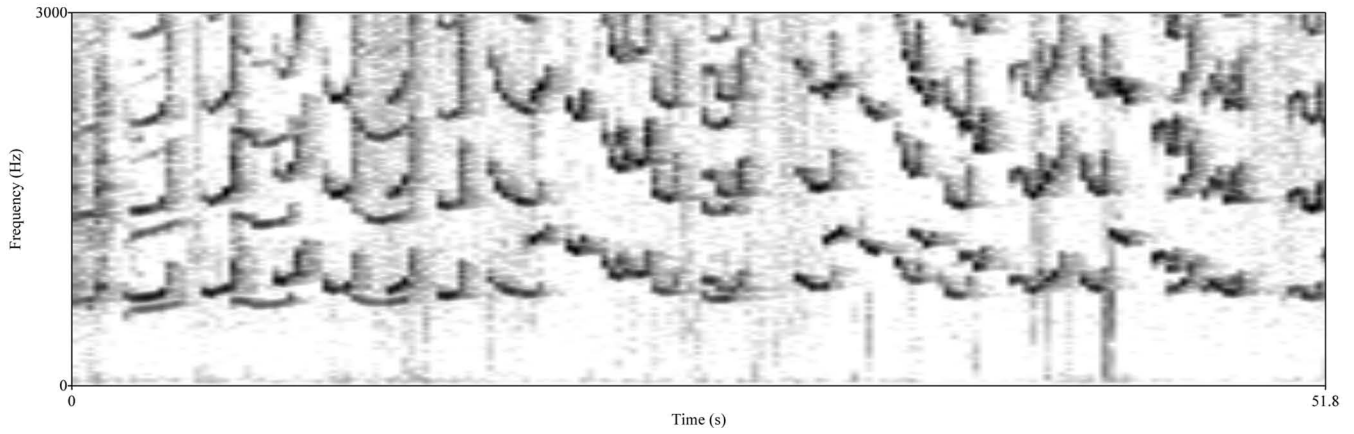
Tab. IV. Effect of sex and context on call parameters.

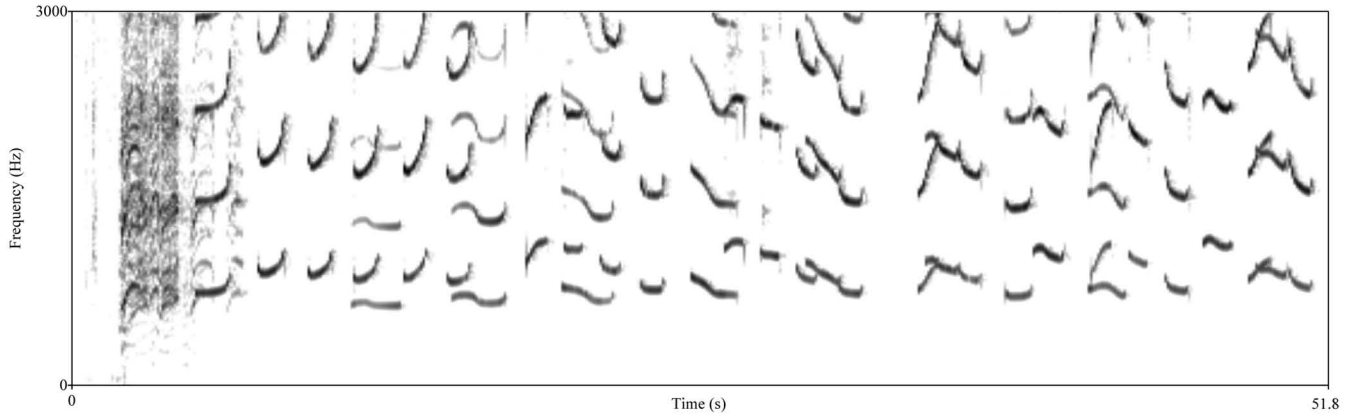|		Roars			Long notes			Descending phrase units		
		df			df			df		
		1; 37	1; 262	1; 262	1; 34	1; 342	1; 342	1; 37	1; 357	1; 357
		Sex	Context	Sex:Context	Sex	Context	Sex:Context	Sex	Context	Sex:Context
Duration	F	0.358	2.730	2.362	99.928	23.312	3.084	53.507	56.065	25.096
	p-value	0.553	0.100	0.126	<0.001	**<0.001**	0.080	<0.001	**<0.001**	<0.001
MeanF0	F	2.100	2.750	0.060	1.570	43.560	3.150	4.600	52.000	6.200
	p-value	0.156	0.098	0.806	0.219	**<0.001**	0.077	0.039	**<0.001**	0.014
MaxF0	F	1.850	4.180	0.530	2.600	23.200	21.400	0.000	26.400	4.300
	p-value	0.183	**0.042**	0.469	0.118	**<0.001**	<0.001	0.843	**<0.001**	0.039
MinF0	F	0.000	1.050	0.860	13.800	43.200	0.700	3.900	36.400	4.400
	p-value	0.976	0.306	0.355	0.001	**<0.001**	0.405	0.393	**<0.001**	0.036
StartF0	F	13.360	5.180	0.220	2.710	17.020	1.110	3.000	24.100	1.600
	p-value	0.001	**0.024**	0.639	0.109	**<0.001**	0.294	0.094	**<0.001**	0.202
EndF0	F	1.280	1.630	0.000	29.500	45.990	17.000	1.000	20.500	5.500
	p-value	0.265	0.202	0.999	<0.001	**<0.001**	<0.001	0.327	**<0.001**	0.019
Pt2Min	F	13.157	0.225	1.880	37.804	0.523	5.424	71.270	0.420	1.780
	p-value	0.001	0.636	0.172	<0.001	0.470	0.020	<0.001	0.519	0.183
Pt2Max	F	16.800	0.490	1.650	106.437	1.587	0.066	29.128	0.370	0.471
	p-value	<0.001	0.485	0.200	<0.001	0.209	0.798	<0.001	0.544	0.493
EnstF0	F	0.391	2.803	1.080	2.632	5.748	30.287	6.857	0.001	1.177
	p-value	0.429	0.057	0.300	0.471	**<0.001**	<0.001	0.009	0.451	0.279

Tab. IV Results of the GLMM for the effect of sex, context and the interaction between sex and context on call parameters and degrees of freedom (df) for each factor. Bold denotes statistically significant results for the factor "context" in the GLMM analyses.
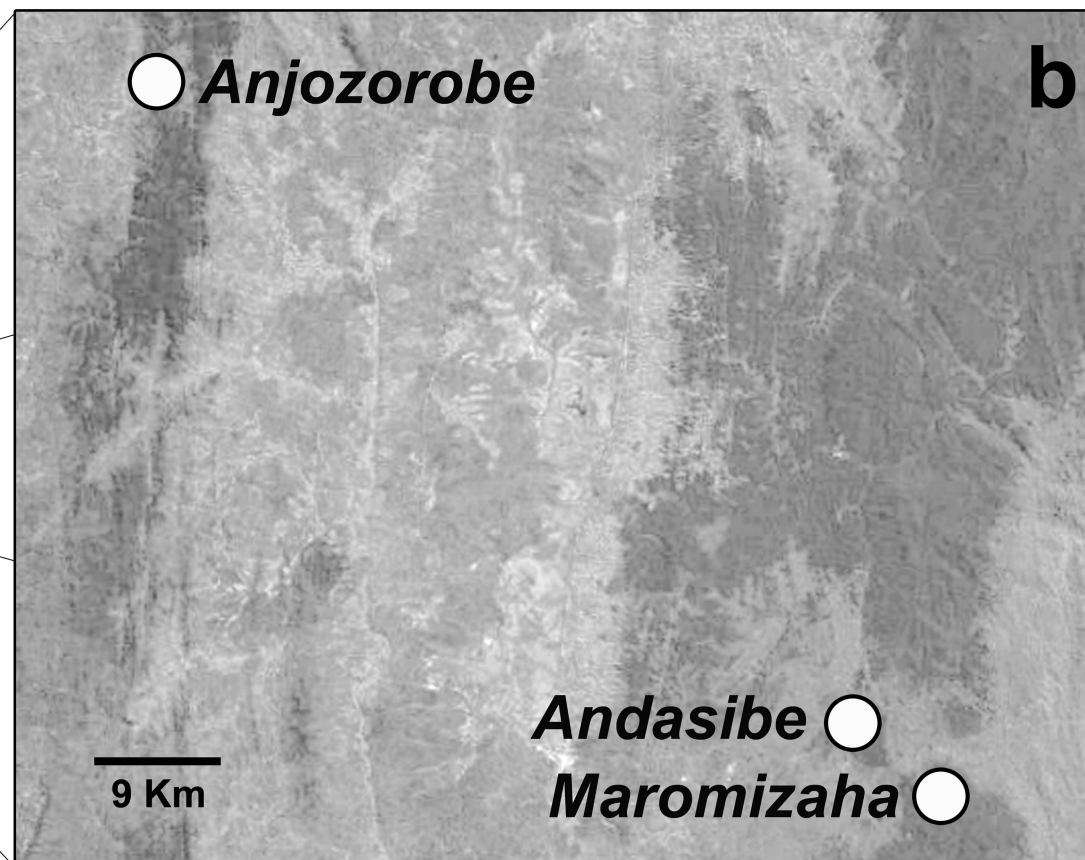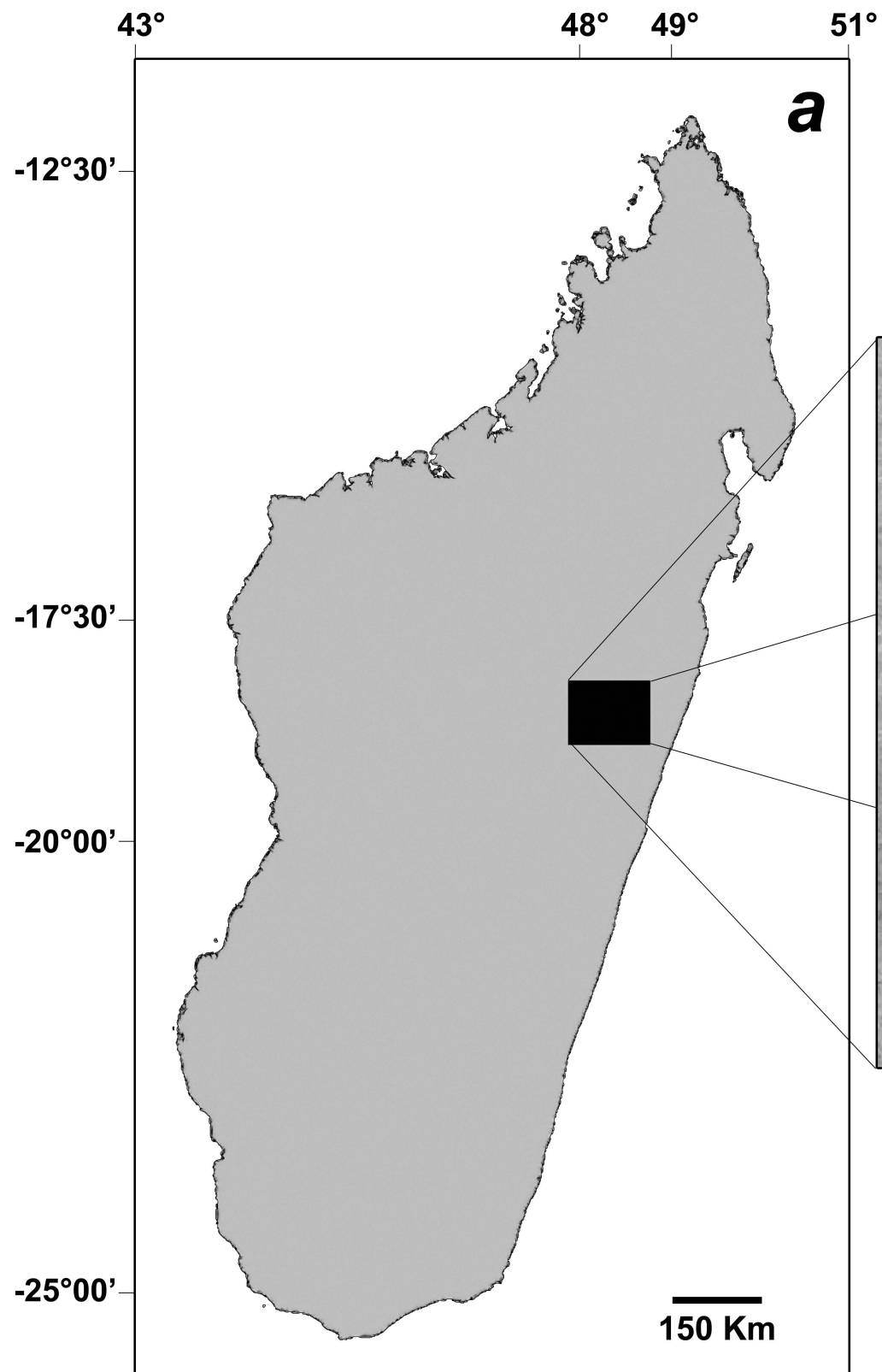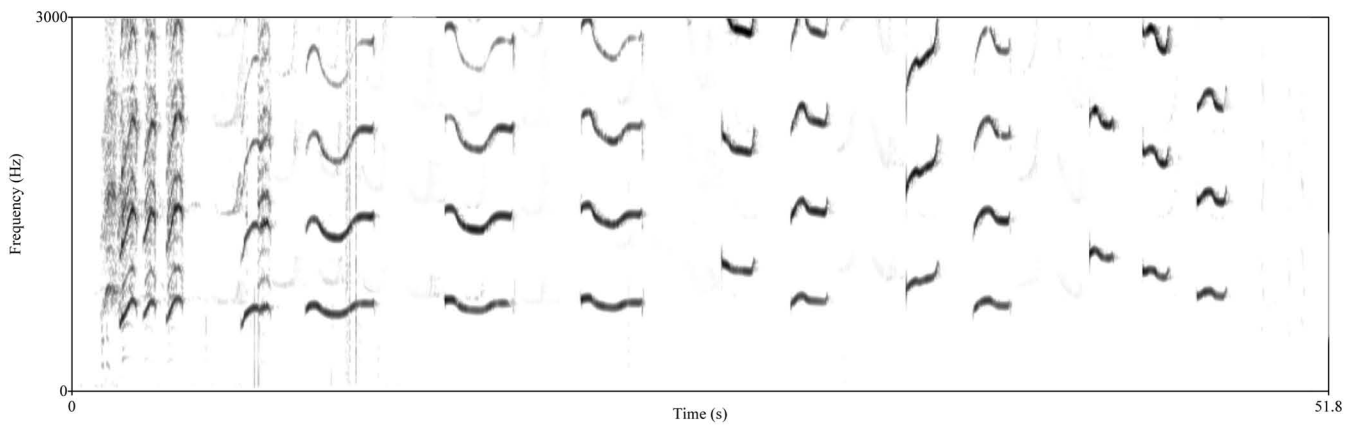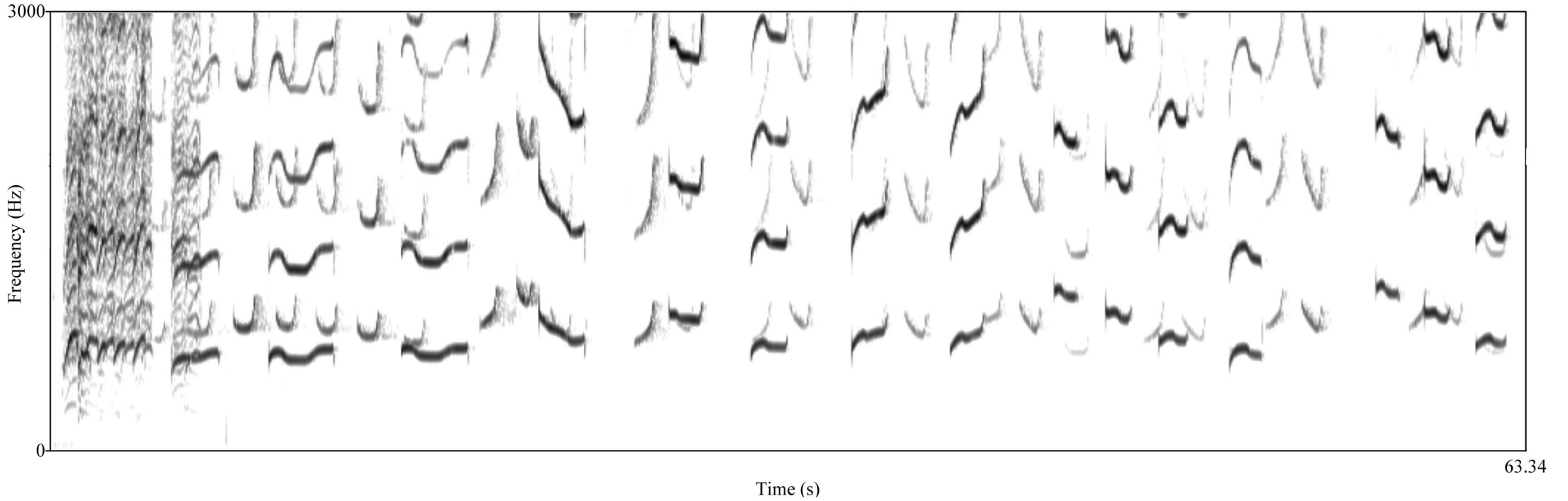
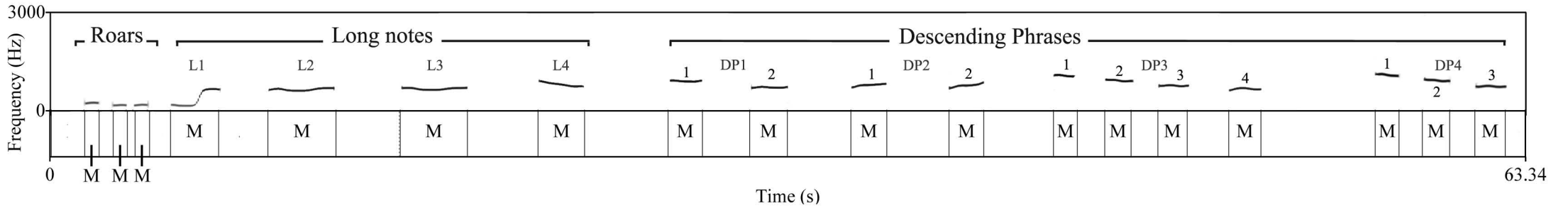Territorial Encounter Song (TS)

Advertisement song (AS)

Cohesion song (CS)

Song

Frequency (Hz)

3000

0

63.34

Time (s)

Adult Male - Individual Contribution

Frequency (Hz)

3000

0

Roars Long notes Descending Phrases

L1 L2 L3 L4 DP1 1 2 DP2 1 2 DP3 1 2 3 4 DP4 1 3 2

M M M M M M M M M M M M M M M M M

0 M M M 63.34

Time (s)

Adult Female - Individual Contribution

Frequency (Hz)

3000

0

Roars Descending Phrases

DP1 1 2 3 4 5 6 DP2 1 2 3 DP3 1 2 DP4 1 2 DP5 1 2 DP6 1 2 3 DP7 1 2 DP8 1 2 DP9 1 2 3

F F F F F F F F F F F F F F F F F F F F F F F F F F F

0 F F F F 63.34

Time (s)