4

5 **An evaluation of the efficiency of passive acoustic monitoring in detecting deer and**

6 **primates in comparison with camera traps**

7

8 Hiroto ENARI[a], Haruka S. ENARI[a], Kei OKUDA[b], Tetsuya MARUYAMA[c], Kana N. OKUDA[d]

9

10 [a]Yamagata University, 1-23 Wakabamachi, Tsuruoka, Yamagata, 997-8555 Japan

11 [b]Hiroshima Shudo University, 1-1-1, Ozuka-higashi, Asaminami-ku, Hiroshima 731-3195, Japan

12 [c]Tochigi Prefectural Forest Management Office, 280 Shimo-koikemachi, Utsunomiya, Tochigi

13 321-2105, Japan

14 [d]The United Graduate School of Agricultural Sciences, Iwate University, 3-18-8 Ueda, Morioka,

15 Iwate 020-8550 Japan

16

17 **Corresponding author**; Hiroto ENARI

18 Tel: +81 235 28 2925   Fax: +81 235 28 2925

19 E-mail: enari@tds1.tr.yamagata-u.ac.jp

Abstract

In recent years, camera traps have rapidly become popular for the large-scale monitoring of wildlife distribution and population; however, we should not ignore the uncertainty regarding the reliability of camera-based monitoring by inexperienced data gatherers. This study introduces passive acoustic monitoring (PAM) as an easier technique for monitoring terrestrial mammals that uses the sound cues that they produce. To validate the efficacy of PAM, we quantitatively compared the detection areas and rates between sound cues (from PAM) and visual cues (from camera traps) of two mammals—the sika deer *Cervus nippon* and the Japanese macaque *Macaca fuscata*—across seven study sites in eastern Japan with different population densities. To collect sound cues, we set up multiple autonomous recording units at the sites and continuously recorded ambient sounds, following a pre-determined schedule. The total recording time reached 9,081 hours for deer and 8,235 hours for macaques. We then built sound recognizers to automatically detect eight target call types from the recorded data. To collect visual cues, we also set multiple camera traps at the same sites and for the same observation periods. The key findings were as follows: (1) the fully automated procedures that only used the recognizers to detect sound cues produced numerous false positive detections when the call type possessed vocal plasticity and variations; (2) the semi-automated procedures, which included an additional step to validate the automated detections by manual screening exhibited a great improvement in the detectability and recall rates of the half of the target calls, reaching >0.70; (3) when using the semi-automated procedures, the frequency of deer and macaque detections per trap-day derived from the sound cues were in most cases approximately dozens of times and several times, respectively, higher than that derived from the visual cues; (4) the main advantage of PAM may be its superior detection areas, which were 100 to 7,000

43   times wider than those of camera traps; and (5) the current success of the recognition of

44   different call types of each species could broaden the use of PAM, which is not possible for

45   camera traps. PAM could provide socio-behavioral data (i.e., the frequencies and types of inter-

46   individual vocal communications) that could help understand the status of population dynamics

47   and the group compositions, in addition to information related to the presence or absence of

48   species.

49

50   **Key words:** *Cervus nippon*; ecoacoustic monitoring; lag-phase management; *Macaca fuscata*;

51   passive acoustic monitoring

52

53   **1 Introduction**

54   Understanding population dynamics of animals responding to ever-changing environments is an

55   important means of assessing whether specific conservation and management strategies are

56   successful (McComb et al., 2010). Population monitoring, especially for mammals with wide

57   home ranges and long life-spans, must be designed at extensive spatiotemporal scales. The

58   continuity of monitoring efforts is key to its success. The participatory approach using camera

59   traps has rapidly become a popular method to ensure continuity (e.g., eMammal program,

60   McShea et al., 2016). The current trend towards low costs, non-invasiveness, and security for

61   both humans and animals has contributed to the popularity of camera traps (Swann and Perkins,

62   2014). Despite their widespread use, attention should also be paid to the bias and dubious

63   reliability of population monitoring with camera traps by inexperienced data gatherers. For

64   example, species identification using photographs might be difficult if species of similar

65   appearance coexist in one area (Meek et al., 2013). Moreover, camera placement might

66   influence capture rate due to differences in species-specific micro-habitat selections (Mann et

67   al., 2014; Kolowski and Forrester, 2017). Thus, to maintain the accurate and precise data quality

68   required for population monitoring, prior training is needed to avoid inter-researcher biases and

69   to ensure a more robust monitoring design (e.g., multiple-classifier approach; Swanson et al.,

70   2016).

71   Despite advances in camera trap techniques, ecoacoustic methods that were historically

72   developed for monitoring cetacean populations in deep waters are now attracting attention for

73   the population monitoring of terrestrial species that are difficult to observe, such as birds in the

74   forest canopy (Blumstein et al., 2011; Marques et al., 2013). Passive acoustic monitoring (PAM)

75   is one of the ecoacoustic methods that is based on the detection of sounds produced by target

76   animals through the use of autonomous recording units (ARUs). The key advantages of PAM

77   over traditional methods are that: (1) sound cues are potentially detected over a larger range

78   than visual cues, not only underwater but also in dense forests (Marques et al., 2013), (2) PAM

79   is applicable to biodiversity assessments since it can simultaneously monitor multiple sound-

80   producing animals (Sueur et al., 2008), and (3) the recent development of recognition

81   algorithms allows automatic detection of target sounds, resulting in considerable efficiency in

82   data processing (Aide et al., 2013; Heinicke et al., 2015). These characteristics ensure the

83   robustness of PAM despite inter-researcher biases and, therefore, ensure that it provides a

84   simple method that can be applied in large-scale monitoring. In spite of its potential, PAM has

85   not been sufficiently validated for monitoring most sound-producing terrestrial mammals: the

86   focus of previous studies has included the forest elephant (*Loxodonta africana cyclotis*;

87   Thompson et al., 2010), the sika deer (*Cervus nippon*; Enari et al., 2017), and some primate

88   species (Heinicke et al., 2015; Kalan et al., 2015; Spillmann et al., 2015; Kalan et al., 2016).

89   In this context, this study aims to validate the PAM technique using surveys on deer and

90   primates, the typical combination of sound-producing mammals that are widely distributed

91   across Asia. In Japan, the sika deer and the Japanese macaque (*Macaca fuscata*) provide an

92   example of this combination. The sika deer is natively distributed across far-east Asia and

93   Russia, but it also occurs in several countries in Europe (including the British Isles), North

94   America, and New Zealand, where they were artificially introduced (Ohdachi et al., 2015;

95   current distribution in eastern Japan is shown in Fig.1 [I]). In recent years, the distribution of the

96   sika deer in native and introduced areas has expanded, resulting in a threat to native vegetation

97   due to their excessive feeding pressure (Takatsuki, 2009). A sensitive method to detect the initial

98   stage of deer invasion, which is mainly led by adult males, is therefore urgently required

99   because conventional techniques, such as spotlight counts, are less useful when searching for

100   deer at low animal densities (Enari et al., 2017).

101   The Japanese macaque is one of the 23 macaque species that inhabit parts of Asia and

102   Northern Africa. It is an endemic species to Japan. Although its distribution is fragmented and it

103   is isolated in eastern Japan due to excessive hunting and forest exploitation in the past, this

104   primate has also been considered an agricultural pest by national policy (Enari and Suzuki,

105   2010). Most primate species, including Japanese macaques, are diurnal and therefore relatively

106   easy to observe. Therefore, most existing methods to evaluate the abundance and distribution of

107   primates strongly depend on direct observations (Plumptre and Cox, 2006; Ross and Reeve,

108   2011). Consequently, alternative methods that do not require the use of sophisticated

109   observation skills have not been sufficiently examined, except for the use of indirect population

110   indices such as nest counts (Plumptre and Cox, 2006), track counts (Enari and Sakamaki, 2011),

111   and camera traps (Pebsworth and LaFleur, 2014).

112   These two mammalian species possess multiple types of calls as an intra- and inter-species

113   communication tool, as is the case with the other most social animals (Seyfarth and Cheney,

114   2010). Therefore, we here validated the efficiency and versatility of PAM to automatically

115   detect each of the different types of calls emitted by them. For the validation, we compared the

116   effective detection areas and rates between sound cues (from PAM) and visual cues (from

117   camera traps) of two terrestrial mammals across seven study sites in eastern Japan with different

118   population densities. In addition, we examined the best monitoring design for PAM by assessing

119   the sampling efficiency across different seasons and within different time slots of the day.

120

121   **2 Methods**

122   **2.1 Features of calls emitted by study animals**

123   Male sika deer tend to gather into single-sex groups throughout the year, except during the

124   rutting season, which extends from late September to early March, and peaks in October (Miura,

125   1984a). During the rut, male groups are dismantled and the dominant male establishes a mating

126   territory with a harem (Miura, 1983). Male deer establishing a mating territory often emit loud

127   calls (Miura, 1984b), similar to most other cervids (Kiley, 1972; Cap et al., 2008). The calls

128   emitted by males of sika deer fall into 10 types (Minami and Kawamichi, 1992). Two of these

129   are loud calls with high frequency, i.e., moans and howls (Fig. 2 [f] and [g]). While the former is

130   generally produced singly and used to retain the tight grouping of females in the harem, the

131   latter is usually repeated several times and emitted to advertise the territory (Miura, 1984b;

132   Minami and Kawamichi, 1992). These behaviors have important implications: while a region

133   where only howl can be heard is in the initial stage of invasion (i.e., the lag phase of population

134   dynamics; Crooks and Soulé, 1999), a region where both moan and howl can be heard is in the

6

135   second stage of invasion (i.e., the early increase phase), since moan suggests female presence.

136   Apart from these two loud calls, the alert bark is also a clearly audible call (Fig. 2 [h]). This is a

137   monosyllabic call and mainly emitted by adult females in the alert situation, for example the

138   sudden appearance of predators (Minami and Kawamichi, 1992).

139   Japanese macaques generally live in troops of 10–100 individuals (Takasaki, 1981), but

140   solitary males or small male groups move around these troops. Non-human primates have

141   acquired diverse call types during the evolution of social behaviors, and of these, Japanese

142   macaques have the widest variety (McComb and Semple, 2005). The calls of Japanese

143   macaques can be heard throughout the year, and fall into six main groups of sounds that serve

144   different socio-ecological functions: (A) affiliative, (B) defensive, (C) aggressive, (D) warning,

145   (E) female estrus, and (F) infant sounds. Overall, these sounds contain 37 different sound

146   subgroups (Itani, 1963). The examples of clear calls heard at a high call rate include coo calls

147   (Fig. 2 [b]; falling within group A, which are typical contact calls that are frequently exchanged

148   among group members to maintain their cohesion during group movement; Koda, 2004), and

149   alarm calls (Fig. 2 [e]; group D), which are defined as being of only a single type, and are

150   related to dogs, snakes and other potential predators (Green, 1975). In addition to these two

151   calls, loud calls (Fig. 2 [a]; group A), screams (Fig. 2 [c]; group B) and aggression (Fig. 2 [d];

152   group C) have been known as typical calls with high-volume sounds.

153

154   **2.2 Study area**

155   For deer monitoring, we selected an area at Nikko mountain range as a typical breeding site

156   with the highest deer population density (50.6 deer/km^2, based on a spotlight survey in 2015,

157   Enari et al., unpublished data). The Aizu was chosen as a site at the early increase phase of the

7

158 deer population, and Asahi, Oguni, Nasu, and Hakkoda were selected as sites in the lag phase

159 (Fig 1 [II]; Table 1). For macaque monitoring, we selected an area at Shirakami as a site where

160 these primates have been continuously present, Asahi as an area where they are currently in the

161 process of population recovery, and Hakkoda as an area where they are rarely observed (Fig. 1

162 [II]).

163   All sites are located in cool-temperate climate zones that experience heavy snowfall. The

164 annual mean ambient temperature and total precipitation during each study year did not vary

165 largely: the highest and lowest values were respectively 12 °C at the Asahi site and 7 °C at the

166 Nikko site, and 2,700 mm at the Oguni site and 1,400 mm at the Aizu site. The forest landscape

167 was also similar between sites, and was typical for cool-temperate zones: broadleaved forests

168 that are mainly composed of beech (*Fagus crenata*), oak (*Quercus crispula*), and elm (*Ulmus*

169 *davidiana*), and conifer plantations patchily distributed, such as cedar (*Cryptomeria japonica*)

170 and larch (*Larix kaempferi*). Regarding the understory vegetation, only low-growing herbaceous

171 plants were observed in the Nikko site due to high levels of deer foraging, whereas dense shrubs

172 and/or bamboo grasses were present at the other sites.

173

174 **2.3 PAM**

175 **2.3.1 Survey design**

176 According to three criteria, call rate, sound volume, and the clarity of social meaning implied in

177 calls, we decided to target three call types for deer (howls, moans, and alert barks) and five call

178 types for macaques (loud calls, coo calls, screams, aggression, and alarm). The basic features of

179 each call type are shown in Fig. 2.

180   We set three to eight ARUs in each site at >1 km intervals to avoid duplicate recordings

181   (Table 1). We mounted all ARUs on trees at a height of 1.5 m, and equipped them with two

182   omnidirectional microphones (flat frequency response between 20 Hz and 20 kHz) pointing

183   horizontally. The gain of the microphone preamplifier was set to the value of +48 dB. We stored

184   the stereo audio files in SDXC (64 GB) or SDHC (128 GB) cards in an uncompressed format

185   (.wav). For the deer monitoring in particular, we activated the ARUs only during dawn and

186   dusk, when the frequency of howls tends to increase (Enari et al., 2017), except for some

187   surveys conducted in the Nikko and Aizu sites where continuous recordings were conducted (for

188   details of time schedules of recording in each survey, see Table 1), because of the following

189   survey (see 2.3.3). Since macaques are typically diurnal mammals, we continuously recorded

190   between sunrise and sunset. Thus, the total recording time reached 9,081 hours (ca. 1.9 terabytes

191   of data) for deer and 8,235 hours (ca. 1.8 terabytes) for macaques.

192

193   **2.3.2 Building recognizers to detect calls**

194   We used Kaleidoscope Pro version 4.5.4 (Wildlife Acoustics Inc., Maynard, MA, USA) to train

195   algorithms (i.e., sound recognizers) to detect particular call types of deer and macaques. The

196   recognition algorithm of this software is based on Hidden Markov models (HMMs), a

197   statistical-state machine model. These methods have been used in sound recognition procedures

198   because of their robustness and flexibility for ecoacoustic signal classification across a variety

199   of species (Kogan and Margoliash, 1998; Ren et al., 2009). We assigned the data sources from

200   the Nikko (period II) and Shirakami sites (autumn), where the population density of the target

201   animals was expected to be the highest surveyed, to the training data, and we built the

202   recognizers of deer and macaque calls in consecutive steps (Fig. 3 [a] I). These training data

203   contained 1,291 clear calls of deer (composed of 926 howls, 356 moans, and 9 alert barks) and

204   840 clear calls of macaques (34 aggression, 13 alarms, 19 coos, 649 loud calls, and 125

205   screams).

206   Step 1: we conducted a clustering analysis to detect and sort similar acoustic signals from

207   all the training data by using Kaleidoscope Pro. For deer, we set the signal parameter to 300–

208   7,000 Hz, 1.0–10.0 seconds duration and 0.35 seconds as the maximum inter-syllables gap. For

209   macaques, we used the settings 50–10,000 Hz, 0.2–10.0 seconds duration and 0.35 seconds as

210   the gap. For the clustering, we set the window of fast Fourier transformation (FFT) at 5.33

211   milliseconds, the maximum number of states for the target size of HMMs at 12, the maximum

212   Euclidean distance to cluster center for building clusters at 0.5, and maximum Euclidean

213   distance from cluster center to include in cluster file outputs at 1.0. These were the default

214   values of the parameter setting of Kaleidoscope Pro. Step 2: we roughly reviewed each of the

215   detections in the clusters through visual and auditory inspection of the sound spectrogram to

216   detect the clusters, including the target calls. Step 3: for all the detections assigned within

217   clusters including the target calls, we manually annotated true positive (i.e., the target calls) and

218   false positive (i.e., the non-target ones) calls, and retrained the initial classifiers for each species

219   by rescanning the same training data using these annotations. The retraining procedure was

220   conducted using the same parameters as step 1, and generated species-specific recognizers. Step

221   4: we validated all the detections derived from the recognizers by manually screening to avoid

222   the possibility that the automated procedure that was reliant on these recognizers would draw

223   false positive detections when the target sounds (i.e., true positives) were rare in a data set

224   (Miller et al., 2012).

225   We then tried to detect respective types of calls for all the remaining data sets based on the

226   above steps, and calculated the mean and 95% confidence interval (CI) of the number of

227   detections, which were estimated by upper and lower 2.5 percentiles calculated from 10,000

228   bootstrap resamples using the package 'simpleboot' in R. 3.4.0 (R Development Core Team,

229   2017). We validated the performance of the automated procedure (i.e., steps 1–3) using the

230   detections manually screened by human experts (i.e., a semi-automated procedure; steps 1–4)

231   (Fig. 3 [a] II). For the validation, we used all the recording data from the Nikko site (period I)

232   for deer and the Asahi site (autumn) for macaques, which were different from the above training

233   data. We then evaluated the quality of the automated procedure by using the multiple indices:

234   accuracy (i.e., the proportion of true positives and true negatives), precision (i.e., the proportion

235   of true positives among the detections predicted by the recognizer), false positive rate (i.e., type-

236   I-error rate), and false negative rate (i.e., type-II-error rate) (Fig. 3 [b]). In addition, we

237   calculated the recall rate (i.e., sensitivity, or the proportion of the target calls that are correctly

238   detected) for both detection procedures. For the calculation of recall rates, we manually counted

239   the number of target calls within all the validation data that was visually and/or audibly

240   detectable in the spectrogram without using the recognizers. For the automated procedure, we

241   also calculated the F-measure (i.e., the harmonic mean of precision and recall) for the model

242   validation.

243

244   **2.3.3 Optimizing sampling schedules**

245   While the seasonal peak of the frequency of howl and moan by deer is clear (i.e., the rut seasons

246   peaking in October; Miura, 1984a), that of macaque calls has not been determined. We therefore

247   compared the detection efficacy across every season with the exception of summer, which was

248   excluded due to the extremely loud ambient noise caused by the buzzing of cicadas (spp. in the

249   Cicadidae family) throughout the recording time. This comparison was conducted at the

11

250 Shirakami site, where the macaque density was expected to be the highest of all sites. Detection

251 efficacy was validated not only by the number of troops detected per trap-day, but also by the

252 frequency and duration of call exchange per troop detection, which could influence the

253 detectability of troops using PAM.

254   To optimize the operating time of ARUs, we also validated the peak time of respective call

255 types during the day using the combined data sources of the Nikko (periods I and II) and Aizu

256 sites (period II) for deer, and the Shirakami and Asahi sites during the autumn for macaques.

257 Here, we tested whether the total number of calls during for each clock time was significantly

258 different among the time periods of the day by comparing the observed and expected values

259 using χ^2 tests with Bonferroni correction.

260

261 **2.3.4 Effective detection distances**

262 We evaluated the effective detection distances of each call type using the same ARUs and

263 recognizers by conducting playback experiments with the use of recorded calls. For the

264 preparation of the experiments, we tried to record each call type using a portable recorder (LS-

265 14, Olympus, Japan) with a directional microphone (AT9944, Audio-Technica, Japan) set to

266 ".wav" format by directly following the target animals, except for deer howl since these data

267 were already collected (Enari et al., 2017). In addition, we measured the sound pressure level

268 (SPL) of calls with a sound level meter (SD-2200, Fuso, Japan), and the distance to the target

269 animal using the range finder (Laser 1000AS, Nikon, Japan) to estimate the theoretical SPL to

270 within 1 m of an audio source (L_1) using the formula: $L_1 = 20\log_{10}(r) + L_2$, where r is the

271 distance to the target animal and L_2 is the SPL at r. Thus, we succeeded in recording aggression,

272 coos, and screams for macaques and measuring those theoretical SPLs from different

273   individuals in November and December 2012 in the Shirakami sites: 86.2 ± 1.1 (standard error,

274   SE) dB for aggression (n = 8), 77.8 ± 2.5 dB for coo (n = 13), and 84.8 ± 1.1 dB for scream (n =

275   25). For the other call types, insufficient samples were available due to difficulties in directly

276   observing the related behaviors in the presence of human observers.

277   We then measured the detection distances of these three macaque call types, as well as

278   howl calls of deer in December 2017 in a flat area adjacent to broadleaf forests at the Shirakami

279   site (ambient temperature = 7.4 °C; background noise = 34.6 dB). We played the recorded calls

280   at the above mean theoretical SPL with a loudspeaker (ER-2830W, TOA, Japan) and recaptured

281   the playbacks from 70−200-m apart at 10-m intervals using the Song Meter SM2+ (Wildlife

282   Acoustics, MA, USA) with sampling rate at 16 kHz. We defined the effective detection distance

283   as the maximum distance at which the signals of each type of call could be automatically

284   detected in audio files using the above recognizers.

285

286   **2.4 Comparing with camera trap surveys**

287   To validate the efficacy of sound cues in comparison to visual cues, we measured the detection

288   rate of the same animals in the same sites by using the camera traps. For this, we measured the

289   relative abundance index, or RAI (i.e., the number of detections per survey effort), of the target

290   species using cameras at every site. For the validation of the particular call types of sika deer,

291   which are highly influenced by the difference in group composition (see section 2.1), we

292   identified the sex of deer for camera trap surveys. We set multiple camera traps around the

293   ARUs at >200-m intervals on suspected animal trails during the study schedules, which

294   overlapped with PAM surveys (Table 1). We used different camera models (the effective

295   detection areas were 65–324 m^2, which were calculated based on the vendor's instructions), but

296   all cameras used were low-glow infrared cameras. To ensure the maximum effective detection

297   area from each camera, we removed the underbrush in front of the cameras. We ensured that we

298   avoided duplicate events when repeatedly capturing the same deer across successive events,

299   judging by their appearance (body size and/or the shape of their antlers). Moreover, we counted

300   serial recording shots of macaques within 1 hour as a single troop emergence event. We

301   calculated the mean and 95% CI in the same manner as previously described (2.3.2). While the

302   current PAM surveys were activated mainly during the peak time of day for the target animal

303   calls, the camera traps operated 24 hours a day. Given this, it should be noted that the detection

304   rate of PAM per trap-day tended to represent an underestimate, compared with that of the

305   camera traps.

306

307   **3 Results**

308   **3.1 Detection performance of PAM**

309   The accuracy and false positive rate of PAM surveys using the automated procedure (i.e., steps

310   1–3) was fairly good for most call types (Table 2). While the precision of PAM showed a good

311   appraisal for every deer call type, it was less precise for every macaque call type, with relatively

312   limited true-positives. The false negative rate and recall rate scored poorly, especially for deer

313   moans and alert barks, and for macaque coos and screams, given that more than half of calls

314   possibly detected were overlooked. These appraisals resulted in poor values of F-measure for

315   every call type except for deer howls and moans. In contrast, the semi-automated procedures

316   (i.e., steps 1–4) showed a remarkable score improvement; the recall rates of the half of call

317   types reached >0.70.

318   When validating the detectability of PAM, controlled trials in all but the Nasu site for both

14

319 deer and macaques demonstrated that the number of detections (per trap-day) derived from the

320 sound cues were several to dozens of times greater than those from the visual cues, or that only

321 the PAM surveys could detect the animals (Table 3).

322   When measuring the effective detection distance of each call type using the recognizers we

323 created here, the longest distance was observed for macaque aggression and deer howls, and

324 both were estimated at 190 m (with a detection area of 11.3 ha). Subsequently, the calls heard at

325 the longest distances were screams (170 m, 9.1 ha) and coos (100 m, 3.1 ha) of macaques. These

326 detection areas were 96 to 1,738 times larger than those available from the currently used

327 camera traps.

328

329 **3.2 Optimal sampling schedule**

330 When evaluating the seasonality of macaque call rates, the number of troops detected per trap-

331 day was highest in autumn, followed by spring and winter (see the Shirakami site in Table 3-b).

332 Similarly, the mean frequency and duration of call exchanges per troop detection by the ARUs

333 in autumn (18.26 ± 10.39 [SE] call counts, 21.27 ± 6.34 minutes, n = 46) was >3 times higher

334 and >2 times longer than those recorded in spring (5.38 ± 3.26 call counts, 10.80 ± 4.48

335 minutes, n = 13). Calls in winter were incalculable due to the small sample size (n = 3).

336   Regarding time variations of the call rate, deer howls and moans showed trimodal

337 distributions peaking at midnight, dawn, and dusk (Fig. 4). Although the frequencies of each

338 macaque call type changed periodically, these all shared a common peak during dusk only.

339

340 **4. Discussion**

341 **4.1 Efficacy of PAM for deer detection**

15

342 The effectiveness of the use of howls as a deer population index between the early and late

343 increase phase of their population had been already confirmed in our pilot study (Enari et al.,

344 2017). The current results are based on wider efforts to enhance the reliability of this finding,

345 not only in the increase phase but also in the lag phase. This could meet the current policy needs

346 for precautionary measures or initial responses to prevent deer population expansion in native

347 (Ohashi et al., 2014) and introduced ranges (Nugent et al., 2011). Moreover, the successful

348 recognition of different call types could broaden the range of PAM usages, which is not possible

349 with camera traps. Since moans and alert barks may only be heard after females begin to appear

350 (see 2.1.1), the detections of these calls could signal whether deer population dynamics have

351 moved into an increase phase triggered by enhancing *allee effects*. Therefore, this signal could

352 be considered as a prime example of information that can only be derived from sound cues,

353 which may also provide socio-behavioral information that could help local authorities to avoid

354 the rapid expansion of deer population, or with *lag-phase management* (Lewis and Kareiva,

355 1993; Crooks and Soulé, 1999; Taylor and Hastings, 2005).

356 Note that sufficient care should be taken when utilizing PAM due to the following three

357 issues. First, it only provides imperfect detection. Neither moans nor howls are generally

358 audible except when adult males establish territories, due to its socio-behavioral meaning

359 (Miura, 1984b; see also 2.1.1). This indicates the possibility that PAM may overlook non-

360 dominant males that exhibit nomadic behavior, as observed in the Nasu site (Table 3). It is

361 known that younger males (i.e., non-territorial males) tend to emit fewer call counts (Miura,

362 1984b; Minami and Kawamichi, 1992).

363 Second, the current automated procedure is limited, as shown in the low recall rates for

364 moans and alert barks (Table 2). This issue might be attributed not only to a deficiency in the

365 training data, but also to specific vocal features: while moans are the call with highest variability

366 in terms of sound frequency and duration, alert barks consist of quite short monosyllabic

367 sounds, and are thus less informative. Hence, an additional manual task is still needed (i.e., step

368 4) to correct this issue. Manually filtering the data, however, is not a daunting task. For

369 example, the amount of time spent executing this additional task in our study was <1 hour per

370 1,000 hours of recording time.

371 Third, deer response to sudden human disturbances may be unpredictable. The Nikko site

372 offered an example of this, where the detection rates varied greatly between the survey years

373 (Table 3). These differences could not be explained by methodological issues inherent in PAM,

374 but was instead due to the actual variation in deer density and social composition over time, as

375 shown in camera trap survey results. In fact, the local government conducted a large-scale cull

376 of deer along a 5 km forest road located in this site, resulting in 17 males and 48 females being

377 shot between surveys. The massive decrease in females directly reduced the frequency of alert

378 barks, which are mainly emitted by females (see 2.1.1), and indirectly halved the frequency of

379 moans emitted by dominant males, due to the disruption of the harems. Aside from those

380 predictable variations in soundscapes, we should carefully note that the frequency of howls

381 tripled between the years despite the reduction in male abundance. The reason for this

382 substantial change was not able to be determined by the current study. However, the possibility

383 that antipredator behavior against external danger factors (i.e., human hunters as a potential

384 predator) would be activated cannot be ruled out, as observed in other Cervidae species such as

385 in the roe deer *Capreolus capreolus* (Reby et al., 1999). Therefore, we should pay particular

386 attention to changes in external environmental conditions when using PAM.

387

4.2 Efficacy of PAM for macaque detection

388

389 Understanding of the effectiveness of sound cues has increased recently, particularly for great

390 apes, which are able to produce very loud sounds; e.g., the evaluations of the group ranging and

391 territory use by using pant hoots and drumming sounds by wild chimpanzees *Pan troglodytes*

392 (Kalan et al., 2016) and the long calls of Bornean orangutans *Pongo pygmaeus wurmbii*

393 (Spillmann et al., 2015). For Japanese macaques, however, previous studies on their vocal

394 communications have focused only on socioecological and behavioral aspects, namely vocal

395 functions, development, and plasticity, leading to the understanding of hominization (Koda and

396 Sugiura, 2010). The current findings support the initial expectation that the sound cues

397 overwhelm the visual cues, even when detecting middle-sized primates such as macaques (Table

398 3). The high detectability rate of PAM could be backed by the high continuity and frequency of

399 vocal communication of macaques during the daytime. Each individual belonging to a troop

400 emits coos as a contact call (see 2.1.2) at 0.4–1.0 times per minute, and the frequency of these

401 calls tends to increase in lower visibility habitats (Koda et al., 2008), where the detectability of

402 camera traps is usually limited (Hofmeester et al., 2017).

403 Similar to the case of sika deer, a major issue that remains unaddressed is the improvement

404 of detection accuracy of macaque calls using the automated procedures, as shown in the high

405 type II error rate and the low recall rate, especially for the coo call. Aside from the shortage of

406 the training data, detection accuracy was affected by the acoustic feature of coo calls—this call

407 is an important contact call to maintain group cohesion (Koda, 2004) and, therefore, possesses

408 vocal plasticity and dialect variations to ensure the success of vocal transmission. In fact, the

409 acoustic features of coo calls vary with background noise levels (Koda and Sugiura, 2010), the

410 distance from other group members (Sugiura, 2007), and the differences in habitat environments

18

411   producing specific patterns of acoustical attenuation (Sugiura et al., 2006; Koda et al., 2008).

412   Although such sound variations could be recognized by human ears (i.e., step 4), results still had

413   detection errors elicited after using HMMs. However, given that the amount of time spent

414   conducting the manual task of step 4 was <2 hours per 1,000 hours of recording time in the

415   current study, it is highly likely that the semi-automated procedure presents an alternative,

416   feasible option for macaque detection.

417

418   **4.3 More effective use of PAM**

419   When conducting population monitoring at wider spatiotemporal scales, the use of camera traps

420   has been considered to be one of the limited available options (McShea et al., 2016; Swanson et

421   al., 2016). However, inter-observer biases cannot be neglected when dealing with camera-based

422   monitoring by inexperienced volunteers. One cause of these biases could be the limited

423   detection areas of camera traps. In fact, the detection areas available from most camera models

424   range between 15.8 and 324.1 m^2 (Meek et al., 2012), which are approximately 1/100–1/7,000

425   of the detection areas available from the current PAM surveys. Moreover, it should be noted that

426   the detection areas of cameras represent the *maximum* sensor detection areas and are highly

427   sensitive to microtopography and vegetation cover, which directly influence infrared sensor

428   performance and the success in species discrimination. Given this, PAM is potentially an easier

429   technique for inexperienced volunteers, and therefore allows them to contribute to sustaining

430   large-scale population monitoring efforts, even under snowfall conditions where most camera

431   traps are ineffective.

432   Note that the current PAM has two major technological challenges to overcome. The first is

433   battery drain, since while the most camera traps operate only when the infrared sensor is

19

434   activated, PAM is continuously operating according to a pre-defined schedule. Thus, PAM

435   surveys should optimize the operating schedule to provide an increase in fieldwork efficiency.

436   For deer monitoring, there may be two possible options: (1) recording only during the peak

437   period (i.e., 3 AM–7 AM and 4 PM–6 PM), which is expected to record 40% of the total deer

438   calls in the day (taken from results in Fig. 4); (2) continuously recoding during times other than

439   the off-peak period (i.e., 7 AM–3 PM), which is expected to record 90% of the calls. For

440   macaque monitoring, while the current findings clearly indicated that autumn (i.e., breeding

441   season) is the best season for PAM (Table 2), the optimal recording time of a day was difficult

442   to further narrow down. However, at dusk, when they creep into their sleeping sites, might be a

443   suitable time to carry out recordings since the frequency of calls steadily tends to increase (Fig.

444   4).

445   The second challenge is to correct for the excessive attenuations of sound amplitude, which

446   could directly influence the detection distance of PAM. The excessive attenuation of animal

447   vocalizations has not been sufficiently verified in real fieldworks, but it is considered to be

448   sensitive to the spatial position of sound source—i.e., with a sound source close to the ground

449   all frequencies are more attenuated than at greater heights—and forest landscapes—e.g., the

450   excessive attenuations for frequencies between 1 and 10 kHz is greater in broad-leaved than in

451   coniferous forests (Marten and Marler, 1977). Thus, the data bias regarding the effective

452   detection area of PAM should be adjusted in response to postures and spatial positions when the

453   target species vocalize (especially for arboreal animals such as most primates), surrounding

454   landscape (not only vegetation type, but also trunk density), and season (leafy or non-leafy

455   periods). If these problems are solved, PAM could allow for the measurement of distances to

456   target species, which may contribute to the population density estimation based on the distance

457 sampling theory (Buckland et al., 1993), and the evaluation of habitat use by localizing the

458 target species using sound cues (e.g., the triangulation method with a microphone array; Wilson

459 et al., 2013).

460

461 **Acknowledgements**

468

469 **References**

470 Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R.,

471 2013. Real-time bioacoustics monitoring and automated species identification. PeerJ 1, e103,

472 http://dx.doi.org/10.7717/peerj.103

473 Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J.L.,

474 Krakauer, A.H., Clark, C., Cortopassi, K.A., 2011. Acoustic monitoring in terrestrial

475 environments using microphone arrays: applications, technological considerations and

476 prospectus. J. Appl. Ecol. 48, 758-767, http://dx.doi.org/10.1111/j.1365-2664.2011.01993.x

477 Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., 1993. Distance sampling:

478 estimating abundance of biological populations. Chapman & Hall, New York, USA.

479 Cap, H., Deleporte, P., Joachim, J., Reby, D., 2008. Male vocal behavior and phylogeny

21

480 in deer. Cladistics 24, 917-931, http://dx.doi.org/10.1111/j.1096-0031.2008.00223.x

481 Crooks, J.A., Soulé, M.E., 1999. Lag times in population explosions of invasive species:

482 causes and implications, in: Sandlund, O.T., Schei, J., Viken, A. (Eds.), Invasive species and

483 biodiversity management. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 103-

484 125.

485 Enari, H., Enari, H., Okuda, K., Yoshita, M., Kuno, T., Okuda, K., 2017. Feasibility

486 assessment of active and passive acoustic monitoring of sika deer populations. Ecol. Indicators

487 79, 155-162, http://dx.doi.org/10.1016/j.ecolind.2017.04.004

488 Enari, H., Sakamaki, H., 2011. Estimation of abundance and distribution of Japanese

489 macaques using track-counts in snow. Acta Theriol. 56, 255-265,

490 http://dx.doi.org/10.1007/s13364-011-0025-1

491 Enari, H., Suzuki, T., 2010. Risk of agricultural and property damage associated with the

492 recovery of Japanese monkey populations. Landscape Urban Plann. 97, 83-91,

493 http://dx.doi.org/10.1016/j.landurbplan.2010.04.014

494 Green, S., 1975. Communication by a graded vocal system in Japanese monkeys, in:

495 Rosenblum, L.A. (Ed.), Primate Behavior. Academic Press, New York, pp. 1-102.

496 Heinicke, S., Kalan, A.K., Wagner, O.J., Mundry, R., Lukashevich, H., Kühl, H.S., 2015.

497 Assessing the performance of a semi-automated acoustic monitoring system for primates.

498 Methods in Ecology and Evolution, http://dx.doi.org/10.1111/2041-210X.12384

499 Hofmeester, T.R., Rowcliffe, J.M., Jansen, P.A., Williams, R., Kelly, N., 2017. A simple

500 method for estimating the effective detection distance of camera traps. Remote Sensing in

501 Ecology and Conservation 3, 81-89, http://dx.doi.org/10.1002/rse2.25

502 Itani, J., 1963. Vocal communication of the wild Japanese monkey. Primates 4, 11-66,

503   http://dx.doi.org/https://doi.org/10.1007/BF01659149

504   Kalan, A.K., Mundry, R., Wagner, O.J., Heinicke, S., Boesch, C., Kühl, H.S., 2015.

505   Towards the automated detection and occupancy estimation of primates using passive acoustic

506   monitoring. Ecol. Indicators 54, 217-226, http://dx.doi.org/10.1016/j.ecolind.2015.02.023

507   Kalan, A.K., Piel, A.K., Mundry, R., Wittig, R.M., Boesch, C., Kuhl, H.S., 2016. Passive

508   acoustic monitoring reveals group ranging and territory use: a case study of wild chimpanzees

509   (*Pan troglodytes*). Frontiers in zoology 13, 34, http://dx.doi.org/10.1186/s12983-016-0167-8

510   Kiley, M., 1972. The vocalizations of ungulates, their causation and function. Z.

511   Tierpsychol. 31, 171-222, http://dx.doi.org/10.1111/j.1439-0310.1972.tb01764.x

512   Koda, H., 2004. Flexibility and context-sensitivity during the vocal exchange of coo calls

513   in wild Japanese macaques (*Macaca fuscata yakui*). Behaviour 141, 1279-1296,

514   http://dx.doi.org/10.1163/1568539042729685

515   Koda, H., Shimooka, Y., Sugiura, H., 2008. Effects of caller activity and habitat visibility

516   on contact call rate of wild Japanese macaques (*Macaca fuscata*). Am. J. Primatol. 70, 1055-

517   1063, http://dx.doi.org/10.1002/ajp.20597

518   Koda, H., Sugiura, H., 2010. The ecological design of the affiliative vocal communication

519   style in wild Japanese macaques: behavioral adjustments to social contexts and environments,

520   in: Nakagawa, N., Nakamichi, M., Sugiura, H. (Eds.), The Japanese Macaques. Springer Japan,

521   Tokyo, pp. 165-190.

522   Kogan, J.A., Margoliash, D., 1998. Automated recognition of bird song elements from

523   continuous recordings using dynamic time warping and hidden Markov models: A comparative

524   study. The Journal of the Acoustical Society of America 103, 2185-2196,

525   http://dx.doi.org/http://dx.doi.org/10.1121/1.421364

526   Kolowski, J.M., Forrester, T.D., 2017. Camera trap placement and the potential for bias

527   due to trails and other features. PloS one 12, e0186679,

528   http://dx.doi.org/https://doi.org/10.1371/journal.pone.0186679

529   Lewis, M., Kareiva, P., 1993. Allee dynamics and the spread of invading organisms.

530   Theor. Popul. Biol. 43, 141-158, http://dx.doi.org/10.1006/tpbi.1993.1007

531   Mann, G.K.H., O'Riain, M.J., Parker, D.M., 2014. The road less travelled: assessing

532   variation in mammal detection probabilities with camera traps in a semi-arid biodiversity

533   hotspot. Biodivers. Conserv. 24, 531-545, http://dx.doi.org/10.1007/s10531-014-0834-z

534   Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J.,

535   Harris, D., Tyack, P.L., 2013. Estimating animal population density using passive acoustics.

536   Biological Reviews 88, 287-309, http://dx.doi.org/10.1111/brv.12001

537   Marten, K., Marler, P., 1977. Sound transmission and its significance for animal

538   vocalization: I. Temperate habitats. Behav. Ecol. Sociobiol. 2, 271-290,

539   http://dx.doi.org/10.1007/bf00299740

540   McComb, B., Zuckerberg, B., Vesely, D., Jordan, C., 2010. Monitoring animal

541   populations and their habitats: A practitioner's guide. CRC Press, Frolida, USA.

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

543   primates. Biol. Lett. 1, 381-385, http://dx.doi.org/10.1098/rsbl.2005.0366

544   McShea, W.J., Forrester, T., Costello, R., He, Z., Kays, R., 2016. Volunteer-run cameras

545   as distributed sensors for macrosystem mammal research. Landscape Ecol. 31, 55-66,

546   http://dx.doi.org/10.1007/s10980-015-0262-9

547   Meek, P.D., Ballard, G., Fleming, P.J.S., 2012. An introduction to camera trapping for

548   wildlife surveys in Australia. PestSmart Toolkit Publication, Canberra, Australia.

549   Meek, P.D., Vernes, K., Falzon, G., 2013. On the reliability of expert identification of

550   small-medium sized mammals from camera trap photos. Wildlife Biology in Practice 9,

551   http://dx.doi.org/10.2461/wbp.2013.9.4

552   Miller, D.A., Weir, L.A., McClintock, B.T., Grant, E.H.C., Bailey, L.L., Simons, T.R.,

553   2012. Experimental investigation of false positive errors in auditory species occurrence surveys.

554   Ecol. Appl. 22, 1665-1674, http://dx.doi.org/https://doi.org/10.1890/11-2129.1

555   Minami, M., Kawamichi, T., 1992. Vocal repertoires and classification of the sika deer

556   *Cervus nippon*. Journal of the Mammalogical Society of Japan 17, 71-94,

557   http://dx.doi.org/10.11238/jmammsocjapan.17.71

558   Miura, S., 1983. Grouping behavior of male sika deer in Nara park, Japan. Journal of the

559   Mammalogical Society of Japan 9, 279-284,

560   http://dx.doi.org/10.11238/jmammsocjapan1952.9.279

561   Miura, S., 1984a. Annual cycles of coat changes, antler regrowth, and reproductive

562   behavior of Sika deer in Nara Park, Japan. Journal of the Mammalogical Society of Japan 10, 1-

563   7

564   Miura, S., 1984b. Social behavior and territoriality in male sika deer (*Cervus nippon*

565   Temminck 1838) during the rut. Z. Tierpsychol. 64, 33-73, http://dx.doi.org/10.1111/j.1439-

566   0310.1984.tb00351.x

567   Nugent, G., McShea, W., Parkes, J., Woodley, S., Waithaka, J., Moro, J., Gutierrez, R.,

568   Azorit, C., Guerrero, F.M., Flueck, W., 2011. Policies and management of overabundant deer

569   (native or exotic) in protected areas. Animal Production Science 51, 384-389,

570   http://dx.doi.org/10.1071/AN10288

571   Ohashi, H., Yoshikawa, M., Oono, K., Tanaka, N., Hatase, Y., Murakami, Y., 2014. The

572   impact of sika deer on vegetation in Japan: setting management priorities on a national scale.

573   Environ. Manage., 1-10, http://dx.doi.org/10.1007/s00267-014-0326-7

574   Ohdachi, S.D., Ishibashi, Y., Iwasa, M.A., Saito, T., 2015. The wild mammals of Japan:

575   second edition. Shoukadoh, Kyoto, p. 506.

576   Pebsworth, P.A., LaFleur, M., 2014. Advancing primate research and conservation

577   through the use of camera traps: introduction to the special Issue. Int. J. Primatol. 35, 825-840,

578   http://dx.doi.org/10.1007/s10764-014-9802-4

579   Plumptre, A.J., Cox, D., 2006. Counting primates for conservation: primate surveys in

580   Uganda. Primates 47, 65-73, http://dx.doi.org/10.1007/s10329-005-0146-8

581   R Development Core Team, 2017. R: A language and environment for statistical

582   computing. R Foundation for Statistical Computing, Vienna, Austria.

583   Reby, D., Cargnelutti, B., Hewison, A., 1999. Contexts and possible functions of barking

584   in roe deer. Anim. Behav. 57, 1121-1128, http://dx.doi.org/10.1006/anbe.1998.1056

585   Ren, Y., Johnson, M.T., Clemins, P.J., Darre, M., Glaeser, S.S., Osiejuk, T.S., Out-Nyarko,

586   E., 2009. A framework for bioacoustic vocalization analysis using hidden markov models.

587   Algorithms 2, 1410-1428, http://dx.doi.org/10.3390/a2041410

588   Ross, C., Reeve, N., 2011. Survey and census methods: population distribution and

589   density, in: Setchell, J.M., Curtis, D.J. (Eds.), Field and laboratory methods in primatology—

590   Second edition. Cambridge University Press, Cambridge, pp. 111-131.

591   Seyfarth, R.M., Cheney, D.L., 2010. Production, usage, and comprehension in animal

592   vocalizations. Brain Lang. 115, 92-100, http://dx.doi.org/10.1016/j.bandl.2009.10.003

593   Spillmann, B., van Noordwijk, M.A., Willems, E.P., Mitra Setia, T., Wipfli, U., van

594   Schaik, C.P., 2015. Validation of an acoustic location system to monitor Bornean orangutan

595 (*Pongo pygmaeus wurmbii*) long calls. Am. J. Primatol. 77, 767-776,

596 http://dx.doi.org/10.1002/ajp.22398

597   Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid acoustic survey for

598 biodiversity appraisal. PloS one 3, e4065, http://dx.doi.org/10.1371/journal.pone.0004065

599   Sugiura, H., 2007. Effects of proximity and behavioral context on acoustic variation in

600 the coo calls of Japanese macaques. Am. J. Primatol. 69, 1412-1424,

601 http://dx.doi.org/10.1002/ajp.20447

602   Sugiura, H., Tanaka, T., Masataka, N., 2006. Sound transmission in the habitats of

603 Japanese macaques and its possible effect on population differences in coo calls. Behaviour 143,

604 993-1012, http://dx.doi.org/10.1163/156853906778623617

605   Swann, D.E., Perkins, N., 2014. Camera trapping for animal monitoring and

606 management: a review of applications, in: Meek, P., Fleming, P. (Eds.), Camera trapping:

607 Wildlife management and research. CSIRO Publishing, Collingwood, Australia, pp. 3-11.

608   Swanson, A., Kosmala, M., Lintott, C., Packer, C., 2016. A generalized approach for

609 producing, quantifying, and validating citizen science data from wildlife images. Conserv. Biol.

610 30, 520-531, http://dx.doi.org/10.1111/cobi.12695

611   Takasaki, H., 1981. Troop size, habitat quality, and home range area in Japanese

612 macaques. Behav. Ecol. Sociobiol. 9, 277-281,

613 http://dx.doi.org/https://doi.org/10.1007/BF00299883

614   Takatsuki, S., 2009. Effects of sika deer on vegetation in Japan: a review. Biol. Conserv.

615 142, 1922–1929, http://dx.doi.org/10.1016/j.biocon.2009.02.011

616   Taylor, C.M., Hastings, A., 2005. Allee effects in biological invasions. Ecol. Lett. 8, 895-

617 908, http://dx.doi.org/10.1111/j.1461-0248.2005.00787.x

618   Thompson, M.E., Schwager, S.J., Payne, K.B., 2010. Heard but not seen: an acoustic

619   survey of the African forest elephant population at Kakum Conservation Area, Ghana. Afr. J.

620   Ecol. 48, 224-231, http://dx.doi.org/https://doi.org/10.1111/j.1365-2028.2009.01106.x

621   Wilson, D.R., Battiston, M., Brzustowski, J., Mennill, D.J., 2013. Sound Finder: a new

622   software approach for localizing animals recorded with a microphone array. Bioacoustics 23,

623   99-112, http://dx.doi.org/10.1080/09524622.2013.827588

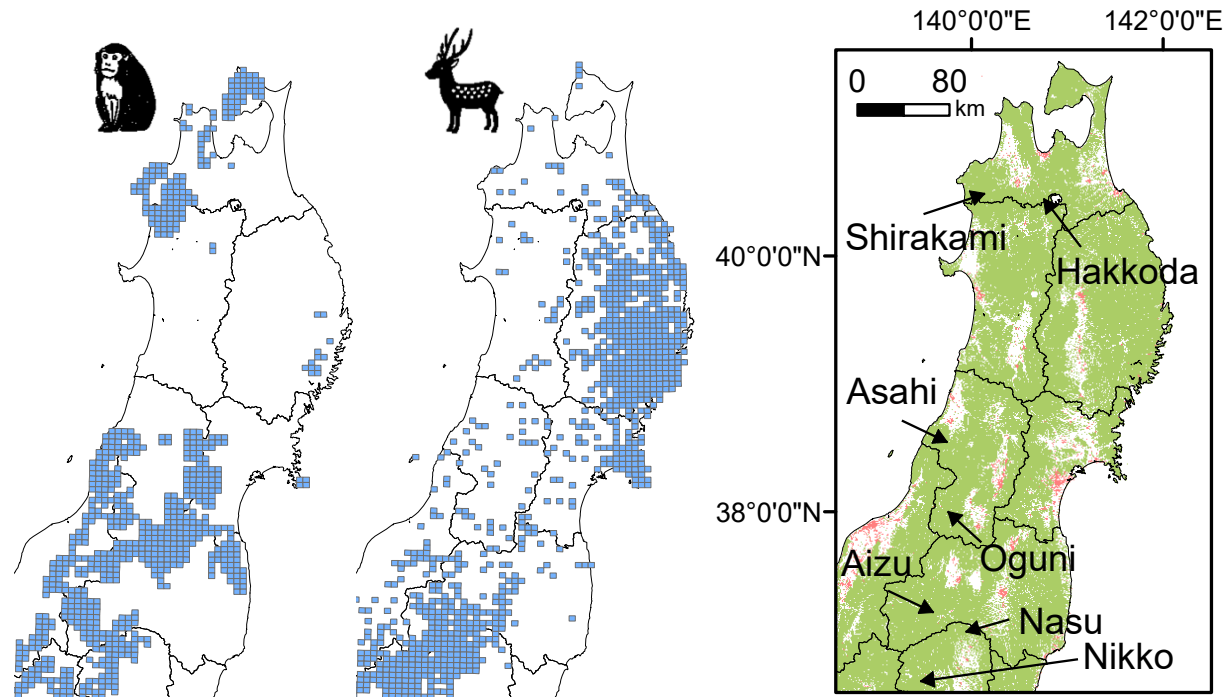624

625

Figure Captions

627

628   Fig. 1. The distributions of Japanese macaques and sika deer (blue grids with 5-km resolution)

629   as of 2014 according to the Ministry of Environment (unpublished data) [I] and geolocations of

630   the study sites [II]. Green- and red-colored areas in [II] show the distribution of woodland and

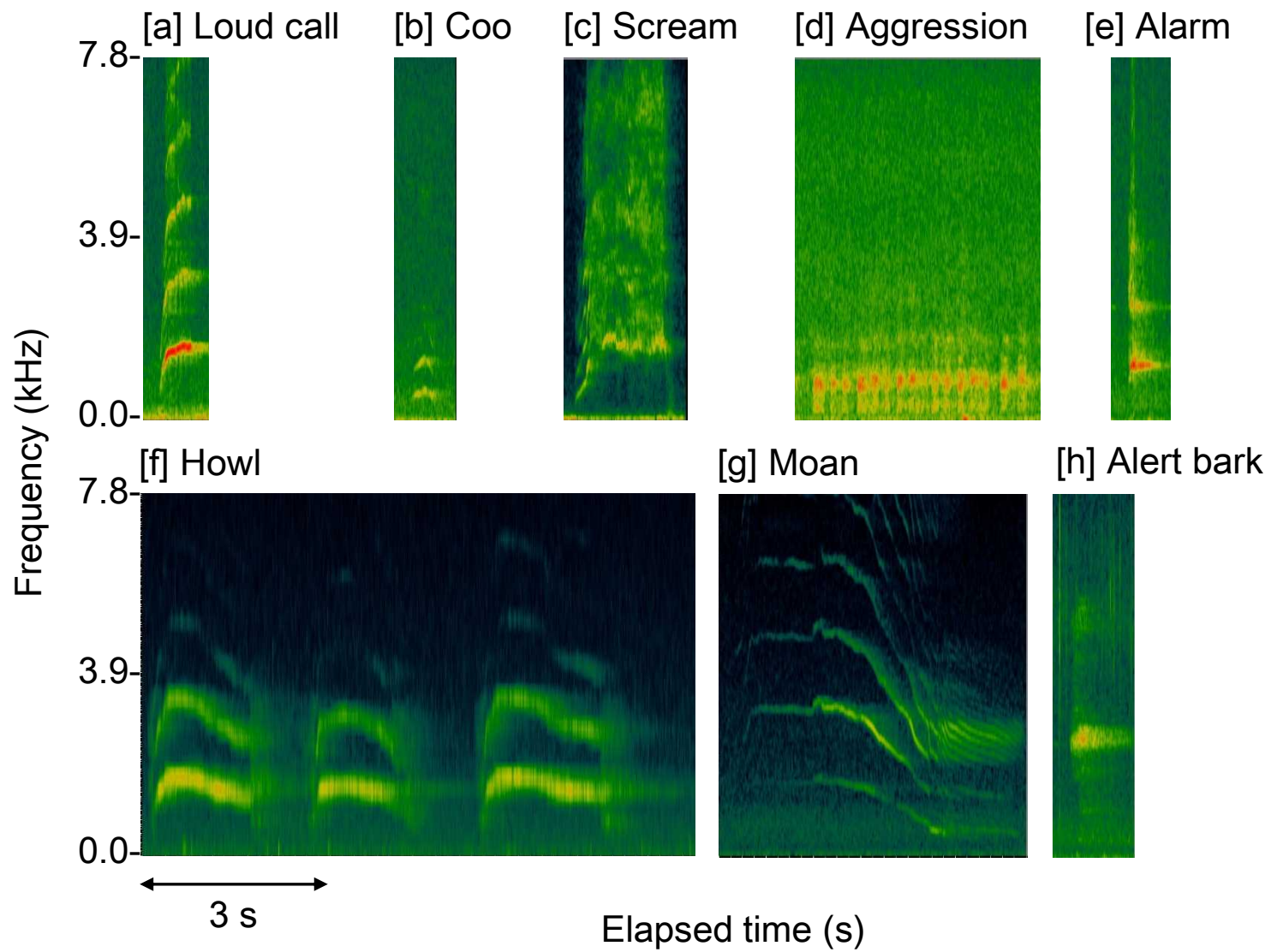631   human residential areas, respectively.

632

633   Fig. 2. Sound spectrograms of calls of Japanese macaques (a–e) and sika deer (f–h), made by

634   using Kaleidoscope Pro ver. 4.5.4 software. The signal becomes stronger from green to red

635   within the spectrograms.

636

637   Fig. 3. Procedures related to automatically/semi-automatically detecting the target type of calls

638   [a] and validating the detections predicted by the sound recognizers with the confusion matrix

639   [b].

640

641   Fig. 4. Hourly variations in the frequency of each type of call emitted by (I) sika deer and (II)

642   Japanese macaques (data sources: Nikko [periods I and II], and Aizu [period II] for deer,

643   Shirakami and Asahi during the autumn for macaques). Deer alert barks and macaque alarms

644   and aggressions were excluded due to the small sample size. Asterisks above the broken lines

645   show the time slots with significantly abundant calls (χ^2 tests with Bonferroni corrections; $\alpha =$

646   0.05). Each call type was manually validated by human experts after being automatically

647   recognized by the models (i.e., semi-automated procedure).

[I] Current distributions of macaque and deer   [II] Placement of study sites

140°0'0"E   142°0'0"E

0   80 km

Shirakami

Hakkoda

40°0'0"N

Asahi

38°0'0"N

Aizu   Oguni

Nasu

Nikko

[a] Loud call [b] Coo [c] Scream [d] Aggression [e] Alarm

[f] Howl [g] Moan [h] Alert bark

Frequency (kHz)

7.8
3.9
0.0

Elapsed time (s)

3 s

[a] Detection of target calls

II. Semi-automated procedure

I. Automated procedure

Step 1: Extract acoustic signals and build initial classifier to sort signals using Hidden Markov models

Step 2: Manually review each cluster and detect the cluster containing the target calls

Step 3: Prepare supervised recording-data by manually annotating all the acoustic signals within the target clusters

Retraining classifier by using supervised data

Sound recognizer

Step 4: Manually screen all the detections from the recognizer

[b] Validation of the detections using the confusion matrix

| | | Detections manually screened by human experts (i.e., procedure II) | |
		Target	Non-target
Detections predicted by recognizer (i.e., procedure I)	Target	True positive (TP)	False positive (FP)
	Non-target	False negative (FN)	True negative (TN)

I. Sika deer

a) Howl

b) Moan

II. Japanese macaque

a) Coo

b) Loud call

c) Scream

d) All repertoires combined

Number of calls per hour

Clock time (Total recording times, h)
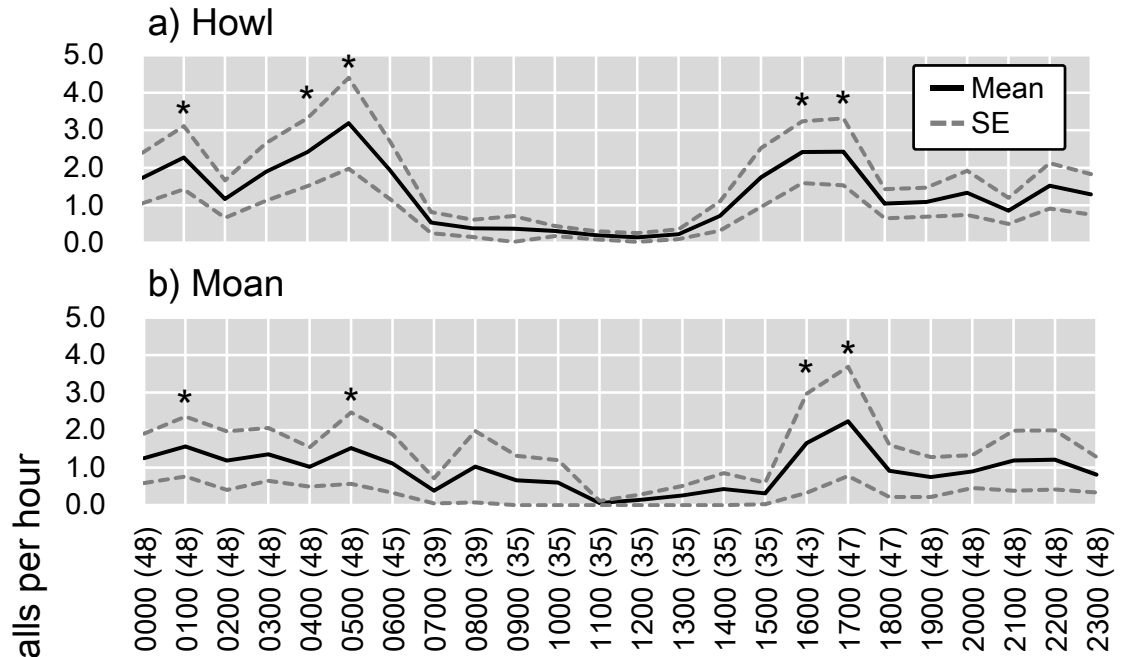
Table 1. Survey designs of PAM and camera trap method for each study site located in eastern Japan
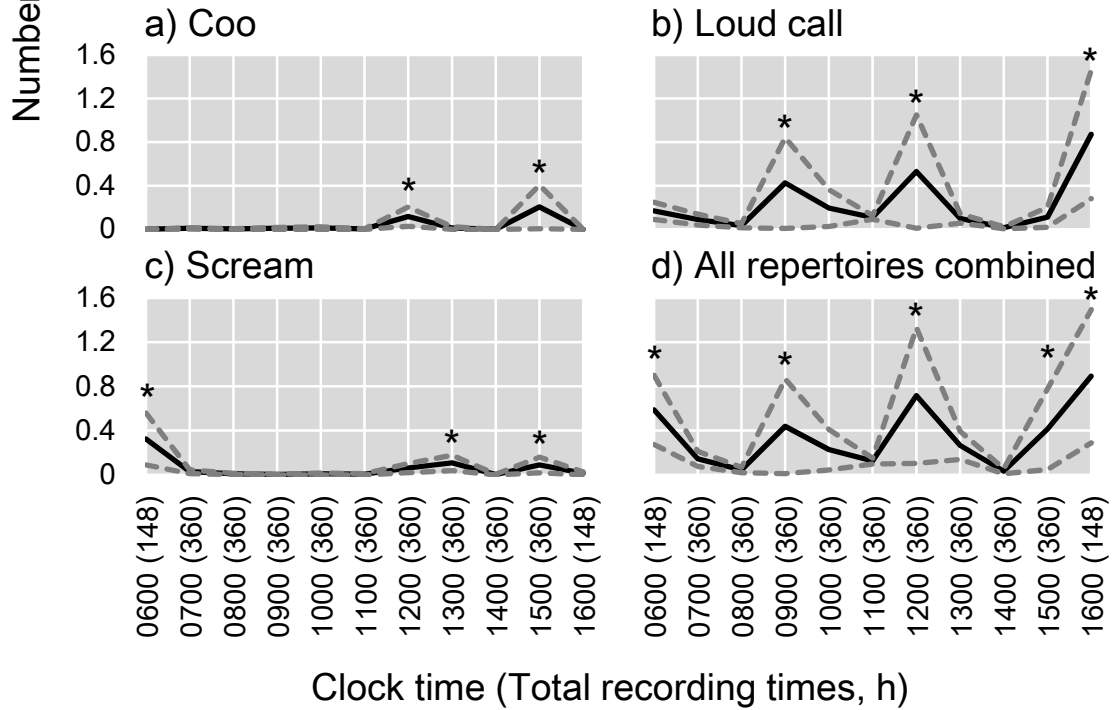
a) Sika deer

	Study site					
	Nikko	Aizu	Asahi	Oguni	Nasu	Hakkoda
PAM						
Model of recorders (ARUs)[a]	SM2+	SM2+	SM2+, SM4	SM2+	SM2+	SM3
#ARUs used	4	5	8	6	5	8
Recording period	I. Oct. 20–22, 2014 II. Oct. 9–14, 2015	I. Sep. 11–Oct. 7, 2016 II. Oct. 7–11, 2016	Oct. 1–Nov. 20, 2017	Oct. 3–Nov. 2, 2016	Sep. 26–Oct. 20, 2017	Sep. 16–Nov. 12, 2017
Time schedule of recording	continuous	I. 2 h both before and after sunrise/sunset II. continuous	2 h both before and after sunrise/sunset	1 h both before and after sunrise/sunset	2 h both before and after sunrise/sunset	1.5 h both before and after sunrise/sunset
Sampling rate (kHz)	16	16	16	16	16	24
Trap-day (trap-h)[c]	I. 12 (157) II. 24 (451)	I. 129 (1,032) II. 25 (425)	364 (2,912)	154 (616)	121 (968)	420 (2,520)
Camera trap method						
Model of cameras[b]	A	B, C	B	B	B	B
#cameras used	9	20	32	28	13	32
Capture delay (min.)	1	1	1	1	1	1
#photos taken per event	1	4	5	5	5	5
Operating period	I. Oct. 1–31, 2014 II. Oct. 1–31, 2015	I. Sep. 11–Oct. 7, 2016 II. Oct. 7–11, 2016	Oct. 1–Nov. 20, 2017	Oct. 3–Nov. 2, 2016	Sep. 26–Oct. 20, 2017	Sep. 16–Nov. 12, 2017
Trap-day[c]	I. 279; II. 279	I. 540; II. 100	1,456	868	325	1,856

b) Japanese macaque

	Study site (Season)				
	Shirakami (spring)	Shirakami (autumn)	Shirakami (winter)	Asahi (autumn)	Hakkoda (spring)
PAM					
Model of ARUs[a]	SM2+	SM2+	SM2+	SM2+	SM3
#ARUs used	7	7	3	7	8
Recording period	Jun. 8–Jul. 1, 2013	Sep. 22–Oct. 13, 2013	Nov. 24–Dec. 14, 2013	Aug. 31–Oct. 2, 2014	Jun. 3–Jul. 3, 2015
Time schedule of recording	6:00–17:00	6:00–17:00	6:30–16:30	7:00–16:00	sunrise–sunset
Sampling rate (kHz)	16	16	16	16	24
Trap-day (trap-h)[c]	147 (1,764)	148 (1,776)	62 (620)	212 (1,908)	223 (2,167)
Camera trap method					
Model of cameras[b]	B	B	B	D	B
#cameras used	21	21	3	28	32
Capture delay (min.)	No delay	No delay	No delay	1	1
#photos taken per event	5	5	5	4	5
Operating period	Jun. 8–Jul. 1, 2013	Sep. 22–Oct. 13, 2013	Nov. 24–Dec. 14, 2013	Aug. 31–Oct. 2, 2014	Jun. 3–Jul. 3, 2015
Trap-day[c]	504	462	63	924	992

[a] All products are made by Wildlife Acoustics, MA, USA.

[b] A: D444 (Moultrie, AL, USA; detection area = ca. 90 m^2), B: HC500 (Reconyx, WI, USA; ca. 324 m^2), C: #119734C (Bushnell, KS, USA; ca. 66 m^2), D: M880 (Moultrie, AL, USA; ca. 65 m^2)

[c] Trap-day or trap-h was calculated as the product of "the number of cameras or recorders" and "effective operating periods (excluding invalid periods caused by dead battery or system error)"

Table 2. Validations of the sound recognizers used to detect particular call types of sika deer and Japanese macaques

	Calculation process[a]	Sika deer			Japanese macaque				
		Howl[b]	Moan	Alert bark	Aggression	Alarm	Coo	Loud call	Scream
Automated procedure									
i. #calls detected	TP	275	390	3	9	1	11	71	16
ii. Accuracy	(TP+TN)/(TP+TN+FP+FN)	0.93	0.86	0.96	0.90	1.00	0.99	0.96	0.96
iii. Precision	TP/(TP+FP)	0.75	0.94	1.00	0.00	0.02	0.03	0.04	0.01
iv. False positive rate	FP/(TN+FP)	0.05	0.02	0.00	0.10	0.00	0.01	0.04	0.05
v. False negative rate	FN/(TP+FN)	0.17	0.41	0.96	0.40	0.50	0.91	0.46	0.78
vi. Recall rate	i/x	0.70	0.43	0.02	0.60	0.50	0.09	0.54	0.22
vii. *F*-measure	2×iii×vi/(iii+vi)	0.72	0.59	0.04	0.00	0.04	0.05	0.07	0.02
Semi-automated procedure									
viii. #calls detected	TP+FN	332	659	85	11	2	113	83	38
ix. Recall rate	viii/x	0.85	0.72	0.52	0.73	0.50	0.90	0.63	0.52
x. #calls manually detected by human experts		392	916	164	15	2	125	131	73
Validation data sources		Nikko (period I)			Asahi (autumn)				

[a] TP = true positive, TN = true negative, FP = false positive, and FN = false negative, which were presented by each procedure

[b] Since "howl" usually involves repetition of the same phrase two to four times, we counted a single set of the continuous calls as one howl.

Table 3. Evaluation of detection rates of particular call types of sika deer and Japanese macaques using PAM, compared with their relative abundances detected by the camera trap method

a) Sika deer

	Study site (period)							
	Nikko (I)	Nikko (II)	Aizu (I)	Aizu (II)	Asahi	Oguni	Nasu	Hakkoda
PAM (semi-automated)								
#howl[a]	332	926	79	86	3	2	0	0
#moan	659	356	0	2	0	0	0	0
#alert bark	85	9	0	0	0	0	0	0
#all repertories combined	89.67	53.79	0.61	3.52	0.01	0.01	0.00	0.00
per trap-day	(12.83–231.83)	(34.00–77.50)	(0.38–0.88)	(1.64–5.80)	(0.00–0.02)	(0.00–0.04)	(0.00–0.00)	(0.00–0.00)
Camera trap								
#deer captured per trap-day[b]	1.34	0.62	0.00	0.02	0.00	0.00	0.01	0.00
	(1.13–1.56)	(0.49–0.77)	(0.00–0.01)	(0.00–0.05)	(0.00–0.01)	(0.00–0.00)	(0.00–0.02)	(0.00–0.00)
#female captured per trap-day[b]	0.62	0.27	0.00	0.00	0.00	0.00	0.00	0.00
	(0.50–0.76)	(0.18–0.36)	(0.00–0.00)	(0.00–0.00)	(0.00–0.00)	(0.00–0.00)	(0.00–0.00)	(0.00–0.00)

b) Japanese macaque

	Study site (Season)				
	Shirakami (spring)	Shirakami (autumn)	Shirakami (winter)	Asahi (autumn)	Hakkoda (spring)
PAM (semi-automated)					
#aggression	3	34	1	11	0
#alarm	0	13	0	2	0
#coo	44	19	0	113	0
#loud call	20	649	3	83	0
#scream	3	125	0	38	2
#all repertories combined per trap-day	0.48 (0.09–1.14)	5.68 (1.32–13.25)	0.06 (0.00–0.15)	1.17 (0.06–2.81)	0.01 (0.00–0.03)
#troops detected per trap-day[c]	0.09 (0.04–0.14)	0.31 (0.22–0.43)	0.05 (0.00–0.11)	0.06 (0.02–0.10)	0.00 (0.00–0.01)
Camera trap					
#troops captured per trap-day[d]	0.01 (0.00–0.02)	0.01 (0.00–0.02)	0.02 (0.00–0.05)	0.01 (0.00–0.02)	0.00 (0.00–0.00)

The values in parentheses show lower and upper 95% confidence intervals, estimated by 10,000 bootstrap resamples

[a] Since "howl" usually involves repetition of the same phrase two to four times, we counted a single set of the continues calls as one howl.

[b] If we repeatedly captured the same deer across successive events, we omitted the duplicate event.

[c] We counted recurring calls recorded within 1 hour as a set of calls generated by a single appearance of a troop.

[d] We counted serial shots of macaques within 1 hour as a single troop emergence event.