

## Original Articles

# An evaluation of the efficiency of passive acoustic monitoring in detecting deer and primates in comparison with camera traps



Hiroto Enari<sup>a,\*</sup>, Haruka S. Enari<sup>a</sup>, Kei Okuda<sup>b</sup>, Tetsuya Maruyama<sup>c</sup>, Kana N. Okuda<sup>d</sup>

<sup>a</sup> Yamagata University, 1-23 Wakabamachi, Tsuruoka, Yamagata 997-8555, Japan

<sup>b</sup> Hiroshima Shudo University, 1-1-1, Ozuka-higashi, Asaminami-ku, Hiroshima 731-3195, Japan

<sup>c</sup> Tochigi Prefectural Forest Management Office, 280 Shimo-koikemachi, Utsunomiya, Tochigi 321-2105, Japan

<sup>d</sup> The United Graduate School of Agricultural Sciences, Iwate University, 3-18-8 Ueda, Morioka, Iwate 020-8550, Japan

## ARTICLE INFO

## Keywords:

*Cervus nippon*

Ecoacoustic monitoring

Lag-phase management

*Macaca fuscata*

Passive acoustic monitoring

## 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 9081 h for deer and 8235 h 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 100–7000 times wider than those of camera traps; and (5) the current success of the recognition of different call types of each species could broaden the use of PAM, which is not possible for camera traps. PAM could provide socio-behavioral data (i.e., the frequencies and types of inter-individual vocal communications) that could help understand the status of population dynamics and the group compositions, in addition to information related to the presence or absence of species.

## 1. Introduction

Understanding population dynamics of animals responding to ever-changing environments is an important means of assessing whether specific conservation and management strategies are successful (McComb et al., 2010). Population monitoring, especially for mammals with wide home ranges and long life-spans, must be designed at extensive spatiotemporal scales. The continuity of monitoring efforts is key to its success. The participatory approach using camera traps has rapidly become a popular method to ensure continuity (e.g., eMammal

program, McShea et al., 2016). The current trend towards low costs, non-invasiveness, and security for both humans and animals has contributed to the popularity of camera traps (Swann and Perkins, 2014). Despite their widespread use, attention should also be paid to the bias and dubious reliability of population monitoring with camera traps by inexperienced data gatherers. For example, species identification using photographs might be difficult if species of similar appearance coexist in one area (Meek et al., 2013). Moreover, camera placement might influence capture rate due to differences in species-specific micro-habitat selections (Mann et al., 2014; Kolowski and Forrester, 2017).

\* Corresponding author.

E-mail address: [enari@tds1.tr.yamagata-u.ac.jp](mailto:enari@tds1.tr.yamagata-u.ac.jp) (H. Enari).

<https://doi.org/10.1016/j.ecolind.2018.11.062>

Received 23 July 2018; Received in revised form 21 November 2018; Accepted 27 November 2018

Available online 04 December 2018

1470-160X/ © 2018 Elsevier Ltd. All rights reserved.

Thus, to maintain the accurate and precise data quality required for population monitoring, prior training is needed to avoid inter-researcher biases and to ensure a more robust monitoring design (e.g., multiple-classifier approach; Swanson et al., 2016).

Despite advances in camera trap techniques, ecoacoustic methods that were historically developed for monitoring cetacean populations in deep waters are now attracting attention for the population monitoring of terrestrial species that are difficult to observe, such as birds in the forest canopy (Blumstein et al., 2011; Marques et al., 2013). Passive acoustic monitoring (PAM) is one of the ecoacoustic methods that is based on the detection of sounds produced by target animals through the use of autonomous recording units (ARUs). The key advantages of PAM over traditional methods are that: (1) sound cues are potentially detected over a larger range than visual cues, not only underwater but also in dense forests (Marques et al., 2013), (2) PAM is applicable to biodiversity assessments since it can simultaneously monitor multiple sound-producing animals (Sueur et al., 2008), and (3) the recent development of recognition algorithms allows automatic detection of target sounds, resulting in considerable efficiency in data processing (Aide et al., 2013; Heinicke et al., 2015). These characteristics ensure the robustness of PAM despite inter-researcher biases and, therefore, ensure that it provides a simple method that can be applied in large-scale monitoring. In spite of its potential, PAM has not been sufficiently validated for monitoring most sound-producing terrestrial mammals: the focus of previous studies has included the forest elephant (*Loxodonta africana cyclotis*; Thompson et al., 2010), the sika deer (*Cervus nippon*; Enari et al., 2017), and some primate species (Heinicke et al., 2015; Kalan et al., 2015; Spillmann et al., 2015; Kalan et al., 2016).

In this context, this study aims to validate the PAM technique using surveys on deer and primates, the typical combination of sound-producing mammals that are widely distributed across Asia. In Japan, the sika deer and the Japanese macaque (*Macaca fuscata*) provide an example of this combination. The sika deer is natively distributed across far-east Asia and Russia, but it also occurs in several countries in Europe (including the British Isles), North America, and New Zealand, where they were artificially introduced (Ohdachi et al., 2015; current distribution in eastern Japan is shown in Fig. 1 [I]). In recent years, the distribution of the sika deer in native and introduced areas has expanded, resulting in a threat to native vegetation due to their excessive feeding pressure (Takatsuki, 2009). A sensitive method to detect the initial stage of deer invasion, which is mainly led by adult males, is therefore urgently required because conventional techniques, such as

spotlight counts, are less useful when searching for deer at low animal densities (Enari et al., 2017).

The Japanese macaque is one of the 23 macaque species that inhabit parts of Asia and Northern Africa. It is an endemic species to Japan. Although its distribution is fragmented and it is isolated in eastern Japan due to excessive hunting and forest exploitation in the past, this primate has also been considered an agricultural pest by national policy (Enari and Suzuki, 2010). Most primate species, including Japanese macaques, are diurnal and therefore relatively easy to observe. Therefore, most existing methods to evaluate the abundance and distribution of primates strongly depend on direct observations (Plumptre and Cox, 2006; Ross and Reeve, 2011). Consequently, alternative methods that do not require the use of sophisticated observation skills have not been sufficiently examined, except for the use of indirect population indices such as nest counts (Plumptre and Cox, 2006), track counts (Enari and Sakamaki, 2011), and camera traps (Pebsworth and LaFleur, 2014).

These two mammalian species possess multiple types of calls as an intra- and inter-species communication tool, as is the case with the other most social animals (Seyfarth and Cheney, 2010). Therefore, we here validated the efficiency and versatility of PAM to automatically detect each of the different types of calls emitted by them. For the validation, we compared the effective detection areas and rates between sound cues (from PAM) and visual cues (from camera traps) of two terrestrial mammals across seven study sites in eastern Japan with different population densities. In addition, we examined the best monitoring design for PAM by assessing the sampling efficiency across different seasons and within different time slots of the day.

## 2. Methods

### 2.1. Features of calls emitted by study animals

Male sika deer tend to gather into single-sex groups throughout the year, except during the rutting season, which extends from late September to early March, and peaks in October (Miura, 1984a). During the rut, male groups are dismantled and the dominant male establishes a mating territory with a harem (Miura, 1983). Male deer establishing a mating territory often emit loud calls (Miura, 1984b), similar to most other cervids (Kiley, 1972; Cap et al., 2008). The calls emitted by males of sika deer fall into 10 types (Minami and Kawamichi, 1992). Two of these are loud calls with high frequency, i.e., moans and howls (Fig. 2 [f] and [g]). While the former is generally

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

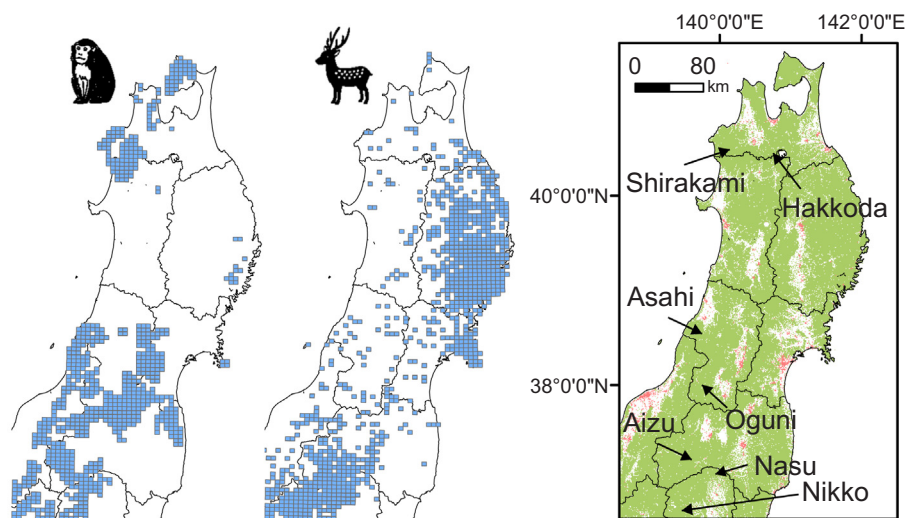
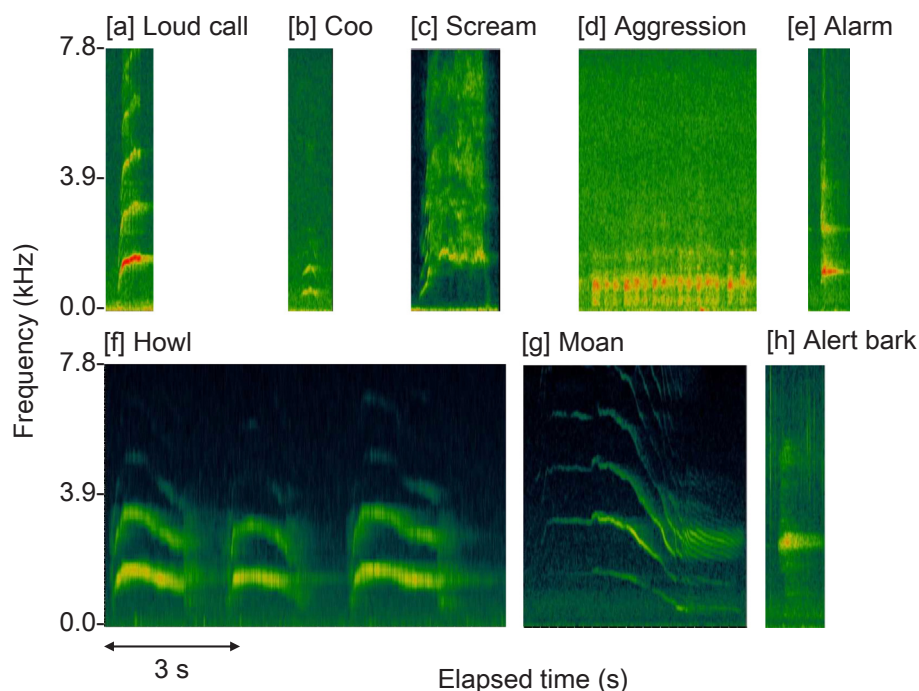


Fig. 1. The distributions of Japanese macaques and sika deer (blue grids with 5-km resolution) as of 2014 according to the Ministry of Environment (unpublished data) [I] and geolocations of the study sites [II]. Green- and red-colored areas in [II] show the distribution of woodland and human residential areas, respectively.



**Fig. 2.** Sound spectrograms of calls of Japanese macaques (a–e) and sika deer (f–h), made by using Kaleidoscope Pro ver. 4.5.4 software. The signal becomes stronger from green to red within the spectrograms.

produced singly and used to retain the tight grouping of females in the harem, the latter is usually repeated several times and emitted to advertise the territory (Miura, 1984b; Minami and Kawamichi, 1992). These behaviors have important implications: while a region where only howl can be heard is in the initial stage of invasion (i.e., the lag phase of population dynamics; Crooks and Soulé, 1999), a region where both moan and howl can be heard is in the second stage of invasion (i.e., the early increase phase), since moan suggests female presence. Apart from these two loud calls, the alert bark is also a clearly audible call (Fig. 2 [h]). This is a monosyllabic call and mainly emitted by adult females in the alert situation, for example the sudden appearance of predators (Minami and Kawamichi, 1992).

Japanese macaques generally live in troops of 10–100 individuals (Takasaki, 1981), but solitary males or small male groups move around these troops. Non-human primates have acquired diverse call types during the evolution of social behaviors, and of these, Japanese macaques have the widest variety (McComb and Semple, 2005). The calls of Japanese macaques can be heard throughout the year, and fall into six main groups of sounds that serve different socio-ecological functions: (A) affiliative, (B) defensive, (C) aggressive, (D) warning, (E) female estrus, and (F) infant sounds. Overall, these sounds contain 37 different sound subgroups (Itani, 1963). The examples of clear calls heard at a high call rate include coo calls (Fig. 2 [b]); falling within group A, which are typical contact calls that are frequently exchanged among group members to maintain their cohesion during group movement; Koda, 2004), and alarm calls (Fig. 2 [e]; group D), which are defined as being of only a single type, and are related to dogs, snakes and other potential predators (Green, 1975). In addition to these two calls, loud calls (Fig. 2 [a]; group A), screams (Fig. 2 [c]; group B) and aggression (Fig. 2 [d]; group C) have been known as typical calls with high-volume sounds.

## 2.2. Study area

For deer monitoring, we selected an area at Nikko mountain range as a typical breeding site with the highest deer population density (50.6 deer/km<sup>2</sup>, based on a spotlight survey in 2015, Enari et al.,

unpublished data). The Aizu was chosen as a site at the early increase phase of the deer population, and Asahi, Oguni, Nasu, and Hakkoda were selected as sites in the lag phase (Fig. 1 [II]; Table 1). For macaque monitoring, we selected an area at Shirakami as a site where these primates have been continuously present, Asahi as an area where they are currently in the process of population recovery, and Hakkoda as an area where they are rarely observed (Fig. 1 [II]).

All sites are located in cool-temperate climate zones that experience heavy snowfall. The annual mean ambient temperature and total precipitation during each study year did not vary largely: the highest and lowest values were respectively 12 °C at the Asahi site and 7 °C at the Nikko site, and 2700 mm at the Oguni site and 1400 mm at the Aizu site. The forest landscape was also similar between sites, and was typical for cool-temperate zones: broadleaved forests that are mainly composed of beech (*Fagus crenata*), oak (*Quercus crispula*), and elm (*Ulmus davidiana*), and conifer plantations patchily distributed, such as cedar (*Cryptomeria japonica*) and larch (*Larix kaempferi*). Regarding the understory vegetation, only low-growing herbaceous plants were observed in the Nikko site due to high levels of deer foraging, whereas dense shrubs and/or bamboo grasses were present at the other sites.

## 2.3. PAM

### 2.3.1. Survey design

According to three criteria, call rate, sound volume, and the clarity of social meaning implied in calls, we decided to target three call types for deer (howls, moans, and alert barks) and five call types for macaques (loud calls, coo calls, screams, aggression, and alarm). The basic features of each call type are shown in Fig. 2.

We set three to eight ARUs in each site at > 1 km intervals to avoid duplicate recordings (Table 1). We mounted all ARUs on trees at a height of 1.5 m, and equipped them with two omnidirectional microphones (flat frequency response between 20 Hz and 20 kHz) pointing horizontally. The gain of the microphone preamplifier was set to the value of +48 dB. We stored the stereo audio files in SDHC (64 GB) or SDXC (128 GB) cards in an uncompressed format (.wav). For the deer monitoring in particular, we activated the ARUs only during dawn and

**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
<i>PAM</i>						
Model of recorders (ARUs) <sup>a</sup>	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) <sup>c</sup>	I. 12 (157) II. 24 (451)	I. 129 (1032) II. 25 (425)	364 (2912)	154 (616)	121 (968)	420 (2520)
<i>Camera trap method</i>						
Model of cameras <sup>b</sup>	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 <sup>c</sup>	I. 279; II. 279	I. 540; II. 100	1456	868	325	1856
b) Japanese macaque						
	Study site (Season)					
	Shirakami (spring)	Shirakami (autumn)	Shirakami (winter)	Asahi (autumn)		Hakkoda (spring)
<i>PAM</i>						
Model of ARUs <sup>a</sup>	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) <sup>c</sup>	147 (1764)	148 (1776)	62 (620)	212 (1908)		223 (2167)
<i>Camera trap method</i>						
Model of cameras <sup>b</sup>	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 <sup>c</sup>	504	462	63	924		992

<sup>a</sup> All products are made by Wildlife Acoustics, MA, USA.

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

<sup>c</sup> 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)”

dusk, when the frequency of howls tends to increase (Enari et al., 2017), except for some surveys conducted in the Nikko and Aizu sites where continuous recordings were conducted (for details of time schedules of recording in each survey, see Table 1), because of the following survey (see 2.3.3). Since macaques are typically diurnal mammals, we continuously recorded between sunrise and sunset. Thus, the total recording time reached 9081 h (ca. 1.9 terabytes of data) for deer and 8235 h (ca. 1.8 terabytes) for macaques.

### 2.3.2. Building recognizers to detect calls

We used Kaleidoscope Pro version 4.5.4 (Wildlife Acoustics Inc., Maynard, MA, USA) to train algorithms (i.e., sound recognizers) to detect particular call types of deer and macaques. The recognition algorithm of this software is based on Hidden Markov models (HMMs), a statistical-state machine model. These methods have been used in sound recognition procedures because of their robustness and flexibility for ecoacoustic signal classification across a variety of species (Kogan

and Margoliash, 1998; Ren et al., 2009). We assigned the data sources from the Nikko (period II) and Shirakami sites (autumn), where the population density of the target animals was expected to be the highest surveyed, to the training data, and we built the recognizers of deer and macaque calls in consecutive steps (Fig. 3 [a] I). These training data contained 1291 clear calls of deer (composed of 926 howls, 356 moans, and 9 alert barks) and 840 clear calls of macaques (34 aggression, 13 alarms, 19 coos, 649 loud calls, and 125 screams).

Step 1: we conducted a clustering analysis to detect and sort similar acoustic signals from all the training data by using Kaleidoscope Pro. For deer, we set the signal parameter to 300–7000 Hz, 1.0–10.0 s duration and 0.35 s as the maximum inter-syllables gap. For macaques, we used the settings 50–10,000 Hz, 0.2–10.0 s duration and 0.35 s as the gap. For the clustering, we set the window of fast Fourier transformation (FFT) at 5.33 ms, the maximum number of states for the target size of HMMs at 12, the maximum Euclidean distance to cluster center for building clusters at 0.5, and maximum Euclidean distance

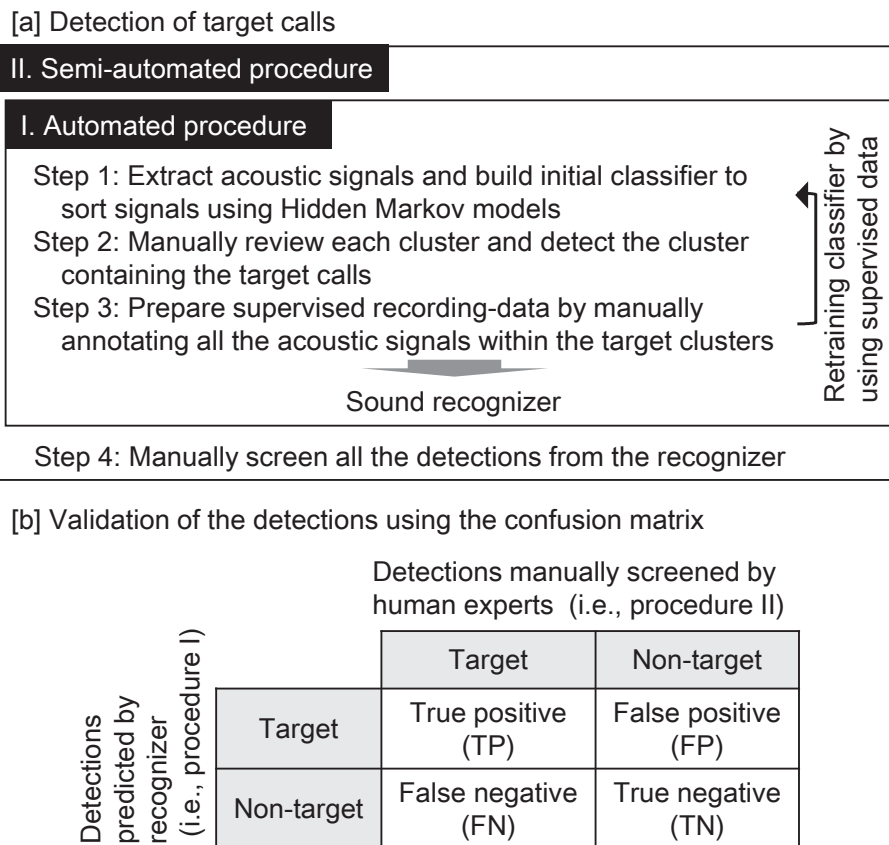


Fig. 3. Procedures related to automatically/semi-automatically detecting the target type of calls [a] and validating the detections predicted by the sound recognizers with the confusion matrix [b].

from cluster center to include in cluster file outputs at 1.0. These were the default values of the parameter setting of Kaleidoscope Pro. Step 2: we roughly reviewed each of the detections in the clusters through visual and auditory inspection of the sound spectrogram to detect the clusters, including the target calls. Step 3: for all the detections assigned within clusters including the target calls, we manually annotated true positive (i.e., the target calls) and false positive (i.e., the non-target ones) calls, and retrained the initial classifiers for each species by re-scanning the same training data using these annotations. The retraining procedure was conducted using the same parameters as step 1, and generated species-specific recognizers. Step 4: we validated all the detections derived from the recognizers by manually screening to avoid the possibility that the automated procedure that was reliant on these recognizers would draw false positive detections when the target sounds (i.e., true positives) were rare in a data set (Miller et al., 2012).

We then tried to detect respective types of calls for all the remaining data sets based on the above steps, and calculated the mean and 95% confidence interval (CI) of the number of detections, which were estimated by upper and lower 2.5 percentiles calculated from 10,000 bootstrap resamples using the package ‘simpleboot’ in R. 3.4.0 (R Development Core Team, 2017). We validated the performance of the automated procedure (i.e., steps 1–3) using the detections manually screened by human experts (i.e., a semi-automated procedure; steps 1–4) (Fig. 3 [a] II). For the validation, we used all the recording data from the Nikko site (period I) for deer and the Asahi site (autumn) for macaques, which were different from the above training data. We then evaluated the quality of the automated procedure by using the multiple indices: accuracy (i.e., the proportion of true positives and true negatives), precision (i.e., the proportion of true positives among the detections predicted by the recognizer), false positive rate (i.e., type-I-error rate), and false negative rate (i.e., type-II-error rate) (Fig. 3 [b]). In addition, we calculated the recall rate (i.e., sensitivity, or the

proportion of the target calls that are correctly detected) for both detection procedures. For the calculation of recall rates, we manually counted the number of target calls within all the validation data that was visually and/or audibly detectable in the spectrogram without using the recognizers. For the automated procedure, we also calculated the *F*-measure (i.e., the harmonic mean of precision and recall) for the model validation.

2.3.3. Optimizing sampling schedules

While the seasonal peak of the frequency of howl and moan by deer is clear (i.e., the rut seasons peaking in October; Miura, 1984a), that of macaque calls has not been determined. We therefore compared the detection efficacy across every season with the exception of summer, which was excluded due to the extremely loud ambient noise caused by the buzzing of cicadas (spp. in the Cicadidae family) throughout the recording time. This comparison was conducted at the Shirakami site, where the macaque density was expected to be the highest of all sites. Detection efficacy was validated not only by the number of troops detected per trap-day, but also by the frequency and duration of call exchange per troop detection, which could influence the detectability of troops using PAM.

To optimize the operating time of ARUs, we also validated the peak time of respective call types during the day using the combined data sources of the Nikko (periods I and II) and Aizu sites (period II) for deer, and the Shirakami and Asahi sites during the autumn for macaques. Here, we tested whether the total number of calls during for each clock time was significantly different among the time periods of the day by comparing the observed and expected values using  $\chi^2$  tests with Bonferroni correction.

2.3.4. Effective detection distances

We evaluated the effective detection distances of each call type

using the same ARUs and recognizers by conducting playback experiments with the use of recorded calls. For the preparation of the experiments, we tried to record each call type using a portable recorder (LS-14, Olympus, Japan) with a directional microphone (AT9944, Audio-Technica, Japan) set to “.wav” format by directly following the target animals, except for deer howl since these data were already collected (Enari et al., 2017). In addition, we measured the sound pressure level (SPL) of calls with a sound level meter (SD-2200, Fuso, Japan), and the distance to the target animal using the range finder (Laser 1000AS, Nikon, Japan) to estimate the theoretical SPL to within 1 m of an audio source ( $L_1$ ) using the formula:  $L_1 = 20\log_{10}(r) + L_2$ , where  $r$  is the distance to the target animal and  $L_2$  is the SPL at  $r$ . Thus, we succeeded in recording aggression, coos, and screams for macaques and measuring those theoretical SPLs from different individuals in November and December 2012 in the Shirakami sites:  $86.2 \pm 1.1$  (standard error, SE) dB for aggression ( $n = 8$ ),  $77.8 \pm 2.5$  dB for coo ( $n = 13$ ), and  $84.8 \pm 1.1$  dB for scream ( $n = 25$ ). For the other call types, insufficient samples were available due to difficulties in directly observing the related behaviors in the presence of human observers.

We then measured the detection distances of these three macaque call types, as well as howl calls of deer in December 2017 in a flat area adjacent to broadleaf forests at the Shirakami site (ambient temperature =  $7.4^\circ\text{C}$ ; background noise = 34.6 dB). We played the recorded calls at the above mean theoretical SPL with a loudspeaker (ER-2830 W, TOA, Japan) and recaptured the playbacks from 70–200-m apart at 10-m intervals using the Song Meter SM2+ (Wildlife Acoustics, MA, USA) with sampling rate at 16 kHz. We defined the effective detection distance as the maximum distance at which the signals of each type of call could be automatically detected in audio files using the above recognizers.

#### 2.4. Comparing with camera trap surveys

To validate the efficacy of sound cues in comparison to visual cues, we measured the detection rate of the same animals in the same sites by using the camera traps. For this, we measured the relative abundance index, or RAI (i.e., the number of detections per survey effort), of the target species using cameras at every site. For the validation of the particular call types of sika deer, which are highly influenced by the difference in group composition (see Section 2.1), we identified the sex of deer for camera trap surveys. We set multiple camera traps around the ARUs at > 200-m intervals on suspected animal trails during the study schedules, which overlapped with PAM surveys (Table 1). We used different camera models (the effective detection areas were 65–324 m<sup>2</sup>, which were calculated based on the vendor’s instructions), but all cameras used were low-glow infrared cameras. To ensure the maximum effective detection area from each camera, we removed the underbrush in front of the cameras. We ensured that we avoided duplicate events when repeatedly capturing the same deer across successive events, judging by their appearance (body size and/or the shape of their antlers). Moreover, we counted serial recording shots of macaques within 1 h as a single troop emergence event. We calculated the mean and 95% CI in the same manner as previously described (2.3.2). While the current PAM surveys were activated mainly during the peak time of day for the target animal calls, the camera traps operated 24 h a day. Given this, it should be noted that the detection rate of PAM per trap-day tended to represent an underestimate, compared with that of the camera traps.

### 3. Results

#### 3.1. Detection performance of PAM

The accuracy and false positive rate of PAM surveys using the automated procedure (i.e., steps 1–3) was fairly good for most call types (Table 2). While the precision of PAM showed a good appraisal for

every deer call type, it was less precise for every macaque call type, with relatively limited true-positives. The false negative rate and recall rate scored poorly, especially for deer moans and alert barks, and for macaque coos and screams, given that more than half of calls possibly detected were overlooked. These appraisals resulted in poor values of  $F$ -measure for every call type except for deer howls and moans. In contrast, the semi-automated procedures (i.e., steps 1–4) showed a remarkable score improvement; the recall rates of the half of call types reached > 0.70.

When validating the detectability of PAM, controlled trials in all but the Nasu site for both deer and macaques demonstrated that the number of detections (per trap-day) derived from the sound cues were several to dozens of times greater than those from the visual cues, or that only the PAM surveys could detect the animals (Table 3).

When measuring the effective detection distance of each call type using the recognizers we created here, the longest distance was observed for macaque aggression and deer howls, and both were estimated at 190 m (with a detection area of 11.3 ha). Subsequently, the calls heard at the longest distances were screams (170 m, 9.1 ha) and coos (100 m, 3.1 ha) of macaques. These detection areas were 96–1738 times larger than those available from the currently used camera traps.

#### 3.2. Optimal sampling schedule

When evaluating the seasonality of macaque call rates, the number of troops detected per trap-day was highest in autumn, followed by spring and winter (see the Shirakami site in Table 3-b). Similarly, the mean frequency and duration of call exchanges per troop detection by the ARUs in autumn ( $18.26 \pm 10.39$  [SE] call counts,  $21.27 \pm 6.34$  min,  $n = 46$ ) was > 3 times higher and > 2 times longer than those recorded in spring ( $5.38 \pm 3.26$  call counts,  $10.80 \pm 4.48$  min,  $n = 13$ ). Calls in winter were incalculable due to the small sample size ( $n = 3$ ).

Regarding time variations of the call rate, deer howls and moans showed trimodal distributions peaking at midnight, dawn, and dusk (Fig. 4). Although the frequencies of each macaque call type changed periodically, these all shared a common peak during dusk only.

### 4. Discussion

#### 4.1. Efficacy of PAM for deer detection

The effectiveness of the use of howls as a deer population index between the early and late increase phase of their population had been already confirmed in our pilot study (Enari et al., 2017). The current results are based on wider efforts to enhance the reliability of this finding, not only in the increase phase but also in the lag phase. This could meet the current policy needs for precautionary measures or initial responses to prevent deer population expansion in native (Ohashi et al., 2014) and introduced ranges (Nugent et al., 2011). Moreover, the successful recognition of different call types could broaden the range of PAM usages, which is not possible with camera traps. Since moans and alert barks may only be heard after females begin to appear (see 2.1), the detections of these calls could signal whether deer population dynamics have moved into an increase phase triggered by enhancing *allee effects*. Therefore, this signal could be considered as a prime example of information that can only be derived from sound cues, which may also provide socio-behavioral information that could help local authorities to avoid the rapid expansion of deer population, or with *lag-phase management* (Lewis and Kareiva, 1993; Crooks and Soulé, 1999; Taylor and Hastings, 2005).

Note that sufficient care should be taken when utilizing PAM due to the following three issues. First, it only provides imperfect detection. Neither moans nor howls are generally audible except when adult males establish territories, due to its socio-behavioral meaning (Miura, 1984b; see also 2.1). This indicates the possibility that PAM may overlook non-

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

Calculation process <sup>a</sup>	Sika deer			Japanese macaque					
	Howl <sup>b</sup>	Moan	Alert bark	Aggression	Alarm	Coo	Loud call	Scream	
<i>Automated procedure</i>									
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
<i>Semi-automated procedure</i>									
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)				

<sup>a</sup> TP = true positive, TN = true negative, FP = false positive, and FN = false negative, which were presented by each procedure.

<sup>b</sup> 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
<i>PAM (semi-automated)</i>								
#howl <sup>a</sup>	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 repertoires combined per trap-day	89.67 (12.83–231.83)	53.79 (34.00–77.50)	0.61 (0.38–0.88)	3.52 (1.64–5.80)	0.01 (0.00–0.02)	0.01 (0.00–0.04)	0.00 (0.00–0.00)	0.00 (0.00–0.00)
<i>Camera trap</i>								
#deer captured per trap-day <sup>b</sup>	1.34 (1.13–1.56)	0.62 (0.49–0.77)	0.00 (0.00–0.01)	0.02 (0.00–0.05)	0.00 (0.00–0.01)	0.00 (0.00–0.00)	0.01 (0.00–0.02)	0.00 (0.00–0.00)
#female captured per trap-day <sup>b</sup>	0.62 (0.50–0.76)	0.27 (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)	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)			
<i>PAM (semi-automated)</i>								
#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 repertoires 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 <sup>c</sup>	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)			
<i>Camera trap</i>								
#troops captured per trap-day <sup>d</sup>	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)			

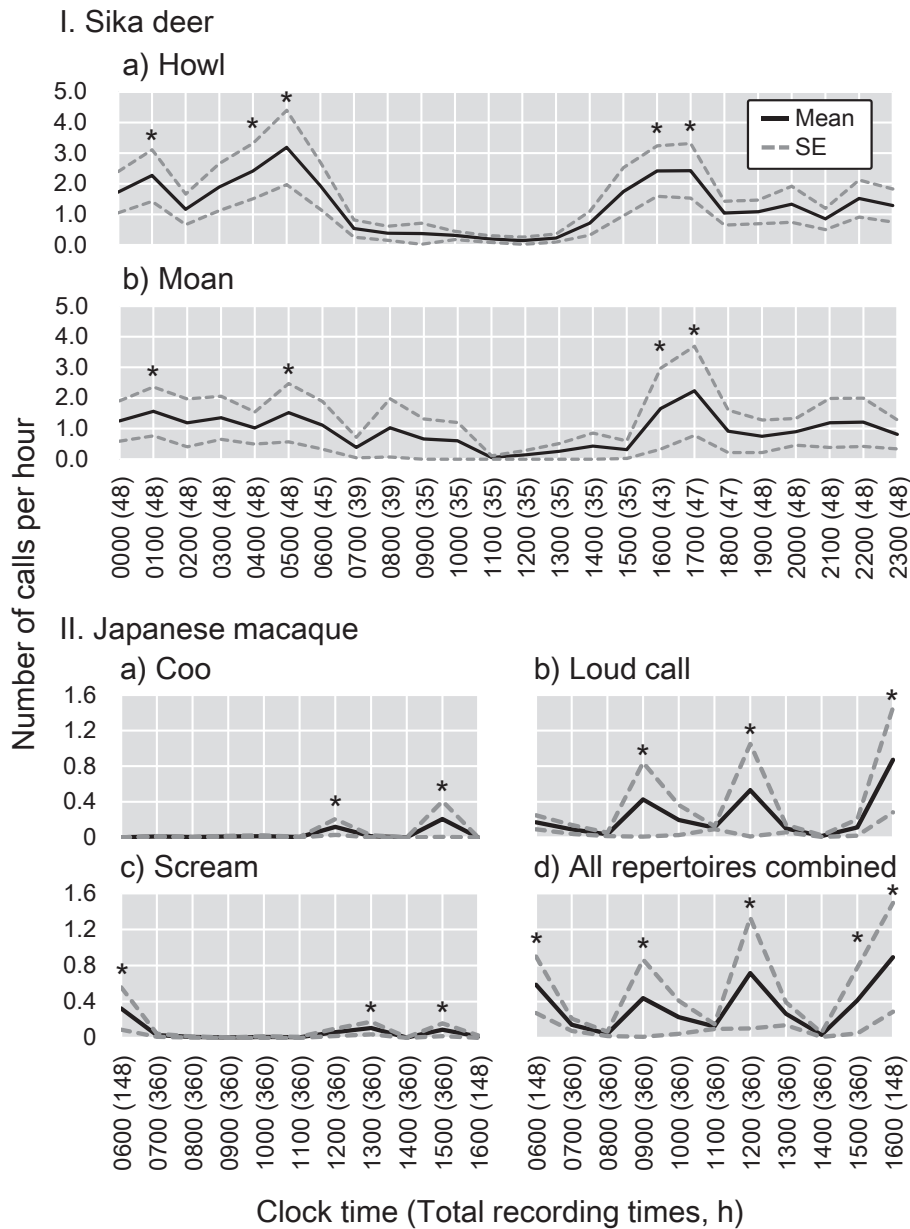
The values in parentheses show lower and upper 95% confidence intervals, estimated by 10,000 bootstrap resamples.

<sup>a</sup> 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.

<sup>b</sup> If we repeatedly captured the same deer across successive events, we omitted the duplicate event.

<sup>c</sup> We counted recurring calls recorded within 1 h as a set of calls generated by a single appearance of a troop.

<sup>d</sup> We counted serial shots of macaques within 1 h as a single troop emergence event.



**Fig. 4.** Hourly variations in the frequency of each type of call emitted by (I) sika deer and (II) Japanese macaques (data sources: Nikko [periods I and II], and Aizu [period II] for deer, Shirakami and Asahi during the autumn for macaques). Deer alert barks and macaque alarms and aggressions were excluded due to the small sample size. Asterisks above the broken lines show the time slots with significantly abundant calls ( $\chi^2$  tests with Bonferroni corrections;  $\alpha = 0.05$ ). Each call type was manually validated by human experts after being automatically recognized by the models (i.e., semi-automated procedure).

dominant males that exhibit nomadic behavior, as observed in the Nasu site (Table 3). It is known that younger males (i.e., non-territorial males) tend to emit fewer call counts (Miura, 1984b; Minami and Kawamichi, 1992).

Second, the current automated procedure is limited, as shown in the low recall rates for moans and alert barks (Table 2). This issue might be attributed not only to a deficiency in the training data, but also to specific vocal features: while moans are the call with highest variability in terms of sound frequency and duration, alert barks consist of quite short monosyllabic sounds, and are thus less informative. Hence, an additional manual task is still needed (i.e., step 4) to correct this issue. Manually filtering the data, however, is not a daunting task. For example, the amount of time spent executing this additional task in our study was < 1 h per 1000 h of recording time.

Third, deer response to sudden human disturbances may be unpredictable. The Nikko site offered an example of this, where the

detection rates varied greatly between the survey years (Table 3). These differences could not be explained by methodological issues inherent in PAM, but was instead due to the actual variation in deer density and social composition over time, as shown in camera trap survey results. In fact, the local government conducted a large-scale cull of deer along a 5 km forest road located in this site, resulting in 17 males and 48 females being shot between surveys. The massive decrease in females directly reduced the frequency of alert barks, which are mainly emitted by females (see 2.1), and indirectly halved the frequency of moans emitted by dominant males, due to the disruption of the harems. Aside from those predictable variations in soundscapes, we should carefully note that the frequency of howls tripled between the years despite the reduction in male abundance. The reason for this substantial change was not able to be determined by the current study. However, the possibility that antipredator behavior against external danger factors (i.e., human hunters as a potential predator) would be activated cannot

be ruled out, as observed in other Cervidae species such as in the roe deer *Capreolus capreolus* (Reby et al., 1999). Therefore, we should pay particular attention to changes in external environmental conditions when using PAM.

#### 4.2. Efficacy of PAM for macaque detection

Understanding of the effectiveness of sound cues has increased recently, particularly for great apes, which are able to produce very loud sounds; e.g., the evaluations of the group ranging and territory use by using pant hoots and drumming sounds by wild chimpanzees *Pan troglodytes* (Kalan et al., 2016) and the long calls of Bornean orangutans *Pongo pygmaeus wurmbii* (Spillmann et al., 2015). For Japanese macaques, however, previous studies on their vocal communications have focused only on socioecological and behavioral aspects, namely vocal functions, development, and plasticity, leading to the understanding of hominization (Koda and Sugiura, 2010). The current findings support the initial expectation that the sound cues overwhelm the visual cues, even when detecting middle-sized primates such as macaques (Table 3). The high detectability rate of PAM could be backed by the high continuity and frequency of vocal communication of macaques during the daytime. Each individual belonging to a troop emits coos as a contact call (see 2.1) at 0.4–1.0 times per minute, and the frequency of these calls tends to increase in lower visibility habitats (Koda et al., 2008), where the detectability of camera traps is usually limited (Hofmeister et al., 2017).

Similar to the case of sika deer, a major issue that remains unaddressed is the improvement of detection accuracy of macaque calls using the automated procedures, as shown in the high type II error rate and the low recall rate, especially for the coo call. Aside from the shortage of the training data, detection accuracy was affected by the acoustic feature of coo calls—this call is an important contact call to maintain group cohesion (Koda, 2004) and, therefore, possesses vocal plasticity and dialect variations to ensure the success of vocal transmission. In fact, the acoustic features of coo calls vary with background noise levels (Koda and Sugiura, 2010), the distance from other group members (Sugiura, 2007), and the differences in habitat environments producing specific patterns of acoustical attenuation (Sugiura et al., 2006; Koda et al., 2008). Although such sound variations could be recognized by human ears (i.e., step 4), results still had detection errors elicited after using HMMs. However, given that the amount of time spent conducting the manual task of step 4 was < 2 h per 1000 h of recording time in the current study, it is highly likely that the semi-automated procedure presents an alternative, feasible option for macaque detection.

#### 4.3. More effective use of PAM

When conducting population monitoring at wider spatiotemporal scales, the use of camera traps has been considered to be one of the limited available options (McShea et al., 2016; Swanson et al., 2016). However, inter-observer biases cannot be neglected when dealing with camera-based monitoring by inexperienced volunteers. One cause of these biases could be the limited detection areas of camera traps. In fact, the detection areas available from most camera models range between 15.8 and 324.1 m<sup>2</sup> (Meek et al., 2012), which are approximately 1/100–1/7000 of the detection areas available from the current PAM surveys. Moreover, it should be noted that the detection areas of cameras represent the *maximum* sensor detection areas and are highly sensitive to microtopography and vegetation cover, which directly influence infrared sensor performance and the success in species discrimination. Given this, PAM is potentially an easier technique for inexperienced volunteers, and therefore allows them to contribute to sustaining large-scale population monitoring efforts, even under snowfall conditions where most camera traps are ineffective.

Note that the current PAM has two major technological challenges

to overcome. The first is battery drain, since while the most camera traps operate only when the infrared sensor is activated, PAM is continuously operating according to a pre-defined schedule. Thus, PAM surveys should optimize the operating schedule to provide an increase in fieldwork efficiency. For deer monitoring, there may be two possible options: (1) recording only during the peak period (i.e., 3 AM–7 AM and 4 PM–6 PM), which is expected to record 40% of the total deer calls in the day (taken from results in Fig. 4); (2) continuously recording during times other than the off-peak period (i.e., 7 AM–3 PM), which is expected to record 90% of the calls. For macaque monitoring, while the current findings clearly indicated that autumn (i.e., breeding season) is the best season for PAM (Table 3), the optimal recording time of a day was difficult to further narrow down. However, at dusk, when they creep into their sleeping sites, might be a suitable time to carry out recordings since the frequency of calls steadily tends to increase (Fig. 4).

The second challenge is to correct for the excessive attenuations of sound amplitude, which could directly influence the detection distance of PAM. The excessive attenuation of animal vocalizations has not been sufficiently verified in real fieldworks, but it is considered to be sensitive to the spatial position of sound source—i.e., with a sound source close to the ground all frequencies are more attenuated than at greater heights—and forest landscapes—e.g., the excessive attenuations for frequencies between 1 and 10 kHz is greater in broad-leaved than in coniferous forests (Marten and Marler, 1977). Thus, the data bias regarding the effective detection area of PAM should be adjusted in response to postures and spatial positions when the target species vocalize (especially for arboreal animals such as most primates), surrounding landscape (not only vegetation type, but also trunk density), and season (leafy or non-leafy periods). If these problems are solved, PAM could allow for the measurement of distances to target species, which may contribute to the population density estimation based on the distance sampling theory (Buckland et al., 1993), and the evaluation of habitat use by localizing the target species using sound cues (e.g., the triangulation method with a microphone array; Wilson et al., 2013).

#### Acknowledgements

We thank Miho Yoshita, Takuya Kuno, Asami Kato, Eri Sugawara, and Masayuki Saito for supporting our data sampling and data processing. All research reported in this paper adhered strictly to the current laws of Japan. This work was supported by JSPS KAKENHI grant number 26701007, YUCOE (C) grant from Yamagata University, Pro Natura Foundation Japan's 28th Pro Natura Fund, Japan, and The Cooperative Research Program of Primate Research Institute, Kyoto University, Japan.

#### References

- Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1, e103. <https://doi.org/10.7717/peerj.103>.
- Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J.L., Krakauer, A.H., Clark, C., Cortopassi, K.A., 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *J. Appl. Ecol.* 48, 758–767. <https://doi.org/10.1111/j.1365-2664.2011.01993.x>.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., 1993. *Distance Sampling: Estimating Abundance of Biological Populations*. Chapman & Hall, New York, USA.
- Cap, H., Deleporte, P., Joachim, J., Reby, D., 2008. Male vocal behavior and phylogeny in deer. *Cladistics* 24, 917–931. <https://doi.org/10.1111/j.1096-0031.2008.00223.x>.
- Crooks, J.A., Soulé, M.E., 1999. Lag times in population explosions of invasive species: causes and implications. In: Sandlund, O.T., Schei, J., Viken, A. (Eds.), *Invasive Species and Biodiversity Management*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 103–125.
- Enari, H., Enari, H., Okuda, K., Yoshita, M., Kuno, T., Okuda, K., 2017. Feasibility assessment of active and passive acoustic monitoring of sika deer populations. *Ecol. Indicators* 79, 155–162. <https://doi.org/10.1016/j.ecolind.2017.04.004>.
- Enari, H., Sakamaki, H., 2011. Estimation of abundance and distribution of Japanese macaques using track-counts in snow. *Acta Theriol.* 56, 255–265. <https://doi.org/10.1007/s13364-011-0025-1>.

- Enari, H., Suzuki, T., 2010. Risk of agricultural and property damage associated with the recovery of Japanese monkey populations. *Landscape Urban Plann.* 97, 83–91. <https://doi.org/10.1016/j.landurbplan.2010.04.014>.
- Green, S., 1975. Communication by a graded vocal system in Japanese monkeys. In: Rosenblum, L.A. (Ed.), *Primate Behavior*. Academic Press, New York, pp. 1–102.
- Heinicke, S., Kalan, A.K., Wagner, O.J., Mundry, R., Lukashevich, H., Kühl, H.S., 2015. Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods Ecol. Evol.* 6, 753–763. <https://doi.org/10.1111/2041-210X.12384>.
- Hofmeister, T.R., Rowcliffe, J.M., Jansen, P.A., Williams, R., Kelly, N., 2017. A simple method for estimating the effective detection distance of camera traps. *Remote Sens. Ecol. Conserv.* 3, 81–89. <https://doi.org/10.1002/rse2.25>.
- Itani, J., 1963. Vocal communication of the wild Japanese monkey. *Primates* 4, 11–66. <https://doi.org/10.1007/BF01659149>.
- Kalan, A.K., Mundry, R., Wagner, O.J., Heinicke, S., Boesch, C., Kühl, H.S., 2015. Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecol. Indicators* 54, 217–226. <https://doi.org/10.1016/j.ecolind.2015.02.023>.
- Kalan, A.K., Piel, A.K., Mundry, R., Wittig, R.M., Boesch, C., Kühl, H.S., 2016. Passive acoustic monitoring reveals group ranging and territory use: a case study of wild chimpanzees (*Pan troglodytes*). *Front. Zool.* 13, 34. <https://doi.org/10.1186/s12983-016-0167-8>.
- Kiley, M., 1972. The vocalizations of ungulates, their causation and function. *Z. Tierpsychol.* 31, 171–222. <https://doi.org/10.1111/j.1439-0310.1972.tb01764.x>.
- Koda, H., 2004. Flexibility and context-sensitivity during the vocal exchange of coo calls in wild Japanese macaques (*Macaca fuscata yakui*). *Behaviour* 141, 1279–1296. <https://doi.org/10.1163/156853904279685>.
- Koda, H., Shimooka, Y., Sugiura, H., 2008. Effects of caller activity and habitat visibility on contact call rate of wild Japanese macaques (*Macaca fuscata*). *Am. J. Primatol.* 70, 1055–1063. <https://doi.org/10.1002/ajp.20597>.
- Koda, H., Sugiura, H., 2010. The ecological design of the affiliative vocal communication style in wild Japanese macaques: behavioral adjustments to social contexts and environments. In: Nakagawa, N., Nakamichi, M., Sugiura, H. (Eds.), *The Japanese Macaques*. Springer Japan, Tokyo, pp. 165–190.
- Kogan, J.A., Margoliash, D., 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. *J. Acoust. Soc. Am.* 103, 2185–2196. <https://doi.org/10.1121/1.421364>.
- Kolowski, J.M., Forrester, T.D., 2017. Camera trap placement and the potential for bias due to trails and other features. *PLoS one* 12, e0186679. <https://doi.org/10.1371/journal.pone.0186679>.
- Lewis, M., Kareiva, P., 1993. Allee dynamics and the spread of invading organisms. *Theor. Popul. Biol.* 43, 141–158. <https://doi.org/10.1006/tpbi.1993.1007>.
- Mann, G.K.H., O'Riain, M.J., Parker, D.M., 2014. The road less travelled: assessing variation in mammal detection probabilities with camera traps in a semi-arid biodiversity hotspot. *Biodivers. Conserv.* 24, 531–545. <https://doi.org/10.1007/s10531-014-0834-z>.
- Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J., Harris, D., Tyack, P.L., 2013. Estimating animal population density using passive acoustics. *Biol. Rev.* 88, 287–309. <https://doi.org/10.1111/brv.12001>.
- Marten, K., Marler, P., 1977. Sound transmission and its significance for animal vocalization: I. Temperate habitats. *Behav. Ecol. Sociobiol.* 2, 271–290. <https://doi.org/10.1007/bf00299740>.
- McComb, B., Zuckerman, B., Vesely, D., Jordan, C., 2010. *Monitoring Animal Populations and Their Habitats: A Practitioner's Guide*. CRC Press, Florida, USA.
- McComb, K., Semple, S., 2005. Coevolution of vocal communication and sociality in primates. *Biol. Lett.* 1, 381–385. <https://doi.org/10.1098/rsbl.2005.0366>.
- McShea, W.J., Forrester, T., Costello, R., He, Z., Kays, R., 2016. Volunteer-run cameras as distributed sensors for macrosystem mammal research. *Landscape Ecol.* 31, 55–66. <https://doi.org/10.1007/s10980-015-0262-9>.
- Meek, P.D., Ballard, G., Fleming, P.J.S., 2012. *An Introduction to Camera Trapping for Wildlife Surveys in Australia*. PestSmart Toolkit Publication, Canberra, Australia.
- Meek, P.D., Vernes, K., Falzon, G., 2013. On the reliability of expert identification of small-medium sized mammals from camera trap photos. *Wildlife Biol. Practice* 9, 1–19. <https://doi.org/10.2461/wbp.2013.9.4>.
- Miller, D.A., Weir, L.A., McClintock, B.T., Grant, E.H.C., Bailey, L.L., Simons, T.R., 2012. Experimental investigation of false positive errors in auditory species occurrence surveys. *Ecol. Appl.* 22, 1665–1674. <https://doi.org/10.1890/11-2129.1>.
- Minami, M., Kawamichi, T., 1992. Vocal repertoires and classification of the sika deer *Cervus nippon*. *J. Mammal. Soc. Jpn.* 17, 71–94. <https://doi.org/10.11238/jmammsojapan.17.71>.
- Miura, S., 1983. Grouping behavior of male sika deer in Nara park, Japan. *J. Mammal. Soc. Jpn.* 9, 279–284. <https://doi.org/10.11238/jmammsojapan1952.9.279>.
- Miura, S., 1984a. Annual cycles of coat changes, antler regrowth, and reproductive behavior of Sika deer in Nara Park, Japan. *J. Mammal. Soc. Jpn.* 10, 1–7.
- Miura, S., 1984b. Social behavior and territoriality in male sika deer (*Cervus nippon Temminck 1838*) during the rut. *Z. Tierpsychol.* 64, 33–73. <https://doi.org/10.1111/j.1439-0310.1984.tb00351.x>.
- Nugent, G., McShea, W., Parkes, J., Woodley, S., Waithaka, J., Moro, J., Gutierrez, R., Azorit, C., Guerrero, F.M., Flueck, W., 2011. Policies and management of over-abundant deer (native or exotic) in protected areas. *Anim. Prod. Sci.* 51, 384–389. <https://doi.org/10.1071/AN10288>.
- Ohashi, H., Yoshikawa, M., Oono, K., Tanaka, N., Hatase, Y., Murakami, Y., 2014. The impact of sika deer on vegetation in Japan: setting management priorities on a national scale. *Environ. Manage.* 54, 631–640. <https://doi.org/10.1007/s00267-014-0326-7>.
- Ohdachi, S.D., Ishibashi, Y., Iwasa, M.A., Saito, T., 2015. *The Wild Mammals of Japan*, second ed. Shoukadoh, Kyoto, pp. 506.
- Pebsworth, P.A., LaFleur, M., 2014. Advancing primate research and conservation through the use of camera traps: introduction to the special Issue. *Int. J. Primatol.* 35, 825–840. <https://doi.org/10.1007/s10764-014-9802-4>.
- Plumptre, A.J., Cox, D., 2006. Counting primates for conservation: primate surveys in Uganda. *Primates* 47, 65–73. <https://doi.org/10.1007/s10329-005-0146-8>.
- R Development Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reby, D., Cargnelutti, B., Hewison, A., 1999. Contexts and possible functions of barking in roe deer. *Anim. Behav.* 57, 1121–1128. <https://doi.org/10.1006/anbe.1998.1056>.
- Ren, Y., Johnson, M.T., Clemins, P.J., Darre, M., Glaeser, S.S., Osiejuk, T.S., Out-Nyarko, E., 2009. A framework for bioacoustic vocalization analysis using hidden markov models. *Algorithms* 2, 1410–1428. <https://doi.org/10.3390/a2041410>.
- Ross, C., Reeve, N., 2011. Survey and census methods: population distribution and density. In: Setchell, J.M., Curtis, D.J. (Eds.), *Field and Laboratory Methods in Primatology—Second Edition*. Cambridge University Press, Cambridge, pp. 111–131.
- Seyfarth, R.M., Cheney, D.L., 2010. Production, usage, and comprehension in animal vocalizations. *Brain Lang.* 115, 92–100. <https://doi.org/10.1016/j.bandl.2009.10.003>.
- Spillmann, B., van Noordwijk, M.A., Willems, E.P., Mitra Setia, T., Wipfli, U., van Schaik, C.P., 2015. Validation of an acoustic location system to monitor Bornean orangutan (*Pongo pygmaeus wurmbii*) long calls. *Am. J. Primatol.* 77, 767–776. <https://doi.org/10.1002/ajp.22398>.
- Sueur, J., Pavoine, S., Hamerlynck, O., Duval, S., 2008. Rapid acoustic survey for biodiversity appraisal. *PLoS One* 3, e4065. <https://doi.org/10.1371/journal.pone.0004065>.
- Sugiura, H., 2007. Effects of proximity and behavioral context on acoustic variation in the coo calls of Japanese macaques. *Am. J. Primatol.* 69, 1412–1424. <https://doi.org/10.1002/ajp.20447>.
- Sugiura, H., Tanaka, T., Masataka, N., 2006. Sound transmission in the habitats of Japanese macaques and its possible effect on population differences in coo calls. *Behaviour* 143, 993–1012. <https://doi.org/10.1163/156853906778623617>.
- Swann, D.E., Perkins, N., 2014. Camera trapping for animal monitoring and management: a review of applications. In: Meek, P., Fleming, P. (Eds.), *Camera Trapping: Wildlife Management and Research*. CSIRO Publishing, Collingwood, Australia, pp. 3–11.
- Swanson, A., Kosmala, M., Lintott, C., Packer, C., 2016. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conserv. Biol.* 30, 520–531. <https://doi.org/10.1111/cobi.12695>.
- Takasaki, H., 1981. Troop size, habitat quality, and home range area in Japanese macaques. *Behav. Ecol. Sociobiol.* 9, 277–281. <https://doi.org/10.1007/BF00299883>.
- Takatsuki, S., 2009. Effects of sika deer on vegetation in Japan: a review. *Biol. Conserv.* 142, 1922–1929. <https://doi.org/10.1016/j.biocon.2009.02.011>.
- Taylor, C.M., Hastings, A., 2005. Allee effects in biological invasions. *Ecol. Lett.* 8, 895–908. <https://doi.org/10.1111/j.1461-0248.2005.00787.x>.
- Thompson, M.E., Schwager, S.J., Payne, K.B., 2010. Heard but not seen: an acoustic survey of the African forest elephant population at Kakum Conservation Area. *Ghana. Afr. J. Ecol.* 48, 224–231. <https://doi.org/10.1111/j.1365-2028.2009.01106.x>.
- Wilson, D.R., Battiston, M., Brzustowski, J., Mennill, D.J., 2013. Sound Finder: a new software approach for localizing animals recorded with a microphone array. *Bioacoustics* 23, 99–112. <https://doi.org/10.1080/09524622.2013.827588>.