Visualizing Soundscape of Animal Vocalizations in Forests Using Robot Audition Techniques

Hao Zhao[1*] Reiji Suzuki[1] Shinji Sumitani[1] Shiho Matsubayashi[2] Takaya Arita[3]
Kazuhiro Nakadai[3, 4] Hiroshi G. Okuno[5, 6]

[1] Nagoya University [2] Osaka University
[3] Tokyo Institute of Technology [4] Honda Research Institute Japan
[5] Kyoto University [6] Waseda University

Abstract: A visualisation of the soundscape dynamics is one of the important topics in ecoacousics. In this paper, We try to use robot audition techniques and ecological methods to visualize the soundscape dynamics of forest animals for a long time. We create two false-color spectrograms based on acoustic indices and direction of arrival of sounds to show the overall dynamics of the soundscape of birds and cicadas in an about four-hour recording in a forest. The preliminary quantitative analysis of their vocal activities also implied that there might exist temporal avoidance behaviors among them.

1 Introduction

Visualization is one of the key techniques when considering roles of sounds in ecoacoustics: the subject to understand their own properties and functions in environments, and the tool for the indirect measurement of biodiversity or habitat quality of environments [1]. Extracting a spatio-temporal structure of a soundscape, which is a combination of sounds that arise from both natural and artificial environments, is essential for both roles in order to track active interactions among individuals and to grasp the overall properties of acoustic events.

We have been proposing and discussing novel applications of robot audition techniques to visualize soundscape dynamics in the directional or spatial domain by using the direction of arrival (DOA) of sound sources obtained from HARKBird, which is a bird song localization software based on a robot audition software HARK (explained later) [2, 3]. Inspired by Towsey et al. [4, 5], we created a false-color spectrogram that visualizes directional (DOA-based) soundscapes in which the color of the spectrogram reflects the direction of arrival of sounds, expecting that we can intuitively recognize directional variations of acoustic events (e.g., different vocalizing individuals or an individual vocalizing in different positions) [2]. We applied this to a 5 min recording with individuals of Zebra Finch, each put in a cage around the microphone array unit, and showed that the extracted visual information can reflect acoustic structures among this simulated group of individuals in the directional domain.

This paper further discusses an application of our framework to a soundscape analysis of a complex situation of vocalizing animals in forests. In particular, we focus on the acoustic dynamics of birds and arthropods, which are major species that dominate the soundscape in forests in early summer. It has been reported that birds are able to adjust both the timing and frequency of their signals to reduce overlap with the signals of other bird species [6, 7], other animals[8] and abiotic noise [9]. Hart et al. showed that birds significantly avoid temporal overlap with cicadas by reducing and often shutting down vocalizations at the onset of cicada signals that utilize the same frequency range [8].

We first illustrate the overall dynamics of the soundscape in about four-hour recording, by showing two false-color spectrograms based on acoustic indices and direction of arrival of sounds. Then, we further illustrate inter- and intra-specific interactions by classifying localized sound sources into bird and cicada vocalizations by making use of a typical acoustic index used in ecoacoustics. The preliminary analysis of their vocalization activities indicated that there might exist temporal overlap avoidance behaviors between birds and cicadas, and intra-specitic turn-taking between cicada individuals.

2 Materials and methods

2.1 HARKBird

HARK is an open-sourced robot audition software consisting of multiple modules for sound source localization, sound source separation, and automatic speech recognition of separated sounds that work on any robot with any microphone configuration [10]. See the website of HARK for detail[1].

HARKBird is a collection of Python scripts that enable us to conduct a field recording using microphone arrays connected to a laptop PC and analyze

*Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601
E-mail:zhao.hao@c.mbox.nagoya-u.ac.jp

[1] https://hark.jp

Figure 1: An experimental field (left) and a recording node (right).

the recording using a network of HARK which are designed to localize and separate bird songs in fields. The HARKBird can estimate the existence and the direction of arrival (DOA) of each sound source by using the MUltiple SIgnal Classification (MUSIC) method [11] based on multiple spectrograms with the short time Fourier transformation. We can further extract separated sounds as wave files for each localized sound using GHDSS (Geometric High order Decorrelation based Source Separation) method. The detailed description of HARKBird and the scripts are available from [12] and our website[2].

2.2 Recording and vocalization localization

We conducted an about 4-hour recording trial in the Inabu field, the experimental forest of Field Science Center, Graduate School of Bioagricultural Sciences, Nagoya University, in central Japan (Fig 1). The forest is mainly composed of conifer plantation (Japanese cedar, Japanese cypress, and red pine), with small patches of broadleaf trees (Quercus, Acer, Carpinus, etc.). In this forest, common bird species are known to vocalize actively during a breeding season.

The recording system is composed of the following components: a server node composed of a single PC; a microphone node (1 (right)) which has a microphone array (TAMAGO-03; System in frontier Inc.) connected with a Rasberry Pi 4; and a Wi-Fi router. The server and the node are connected together by the Wi-Fi, which enables us to control the node remotely. We placed the node in the field where there were some songbirds and cicadas (1 (left)). A recording started at 11:00am, June 27th, 2020 and ended at 3:20pm. In the end, we got thirteen 20-minute recordings with a total duration of four hours and 20 minutes.

We used the HARKBird to export the information about localized sound sources (i.e., the beginning and end time, DOA, and its separated sound file (wave file)). In this paper, we limited the frequency range for sound source localization to 2.5 - 3.5kHz, in order to localize vocalizations of birds and cicadas around this range. This is because some major species of songbirds (Blue-and-white Flycatcher (*Cyanoptila cyanomelana*), Red-billed leiothrix (*Leiothrix lutea*), Eastern-crowned Warbler (*Phylloscopus coronatus*) and Japanese Bush Warbler (*Horornis diphone*)) and some cicadas (*Terpnosia nigricosta*) were singing around the microphone and sharing this frequency range. We adjusted the other parameters in HARKBird to localize these vocalizations as many as possible.

[2] `http://www.alife.cs.is.nagoya-u.ac.jp/~reiji/HARKBird/`

2.3 Soundscape visualization with false-color spectrograms

2.3.1 Acoustic index-based soundscape

Following Towsey et. al. [4], we create a false-color spectrogram based on three acoustic indices: acoustic complexity index (ACI) [13], temporal entropy in frequency bins (H[t]) and acoustic cover (CVR). Each original multi-channel recording (16 bit, 1.6 kHz) is mixed down to a single channel and its amplitude spectrogram (256 frequency bins for 8 kHz, 512 samples for each frame) is created using FFT, which is further divided into 10-second segments. The three types of the spectrum are calculated for each segment as follows:

H[t] spectrum: The temporal entropy of each frequency bin in the amplitude spectrogram. The amplitude values (overtime in a focal frequency bin) are normalized to the unit area and treated as a probability mass function. We calculate Shannon's entropy of this function, which is normalized by the maximum value. This index is useful for picking up infrequent vocalizations.

ACI spectrum: For each frequency bin, we calculate the average absolute fractional change in spectral amplitude from one spectrum to the next [13]. This index is proposed to estimate the abundance of bird vocalizations in a target soundscape.

CVR spectrum: For each frequency bin, we calculate the fraction of values where the spectral power exceeds the noise power (i.e., the average over the values in the frequency bins of 5-8 kHz).

We get a three spectrum matrix of the whole recording with three acoustic indices. Then, we create a false-color image by mapping the values of the three indices to the brightness of the RGB components of each pixel: red=ACI, green=1-H[t] and blue=CVR. The scaled values were assigned to each color in order to make the value differences clearer.

2.3.2 DOA-based soundscape

We create another false-color spectrogram that visualizes DOA-based soundscapes (Fig. 2), proposed in [2], according to the procedures as follows:

1. We generate a grayscale spectrogram of the whole original recording, where the (brighter) grayscale value of each pixel reflects the (higher) energy at the corresponding time and frequency.

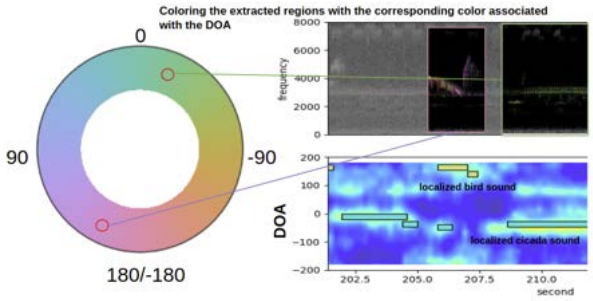

Figure 2: An overview of DOA-based spectrogram. (left) A circular color map, (top right) spectrogram, and (bottom right) MUSIC spectrum (the likelihood of sound existence in the space of time and DOA).

2. We generate a grayscale spectrogram of each separated sound, and extract pixels of which corresponding energy (dB normalized by the maximum value) is higher than 0.9 × the average value over the spectrogram.

3. We pick a color in a circular color map that corresponds to the DOA of each separated sound.

4. We assign the picked color in (3) to those of the pixels of the spectrogram in (1) that correspond to the extracted pixels in (2).

2.4 Classification of bird and cicada vocalizations and their interaction analyses

It is reported that CVR well responds to the continuous cicada chorus [4]. Our preliminary observations of the two spectrograms showed that CVR values around the frequency ranges of vocalizations on which we focus are significantly different between birds (low) and cicadas (high). We classified the localized sound sources into three classes (birds, cicadas, and noise) as follows:

1. We calculate the CVR values of 256 frequency bins of each separated sound file with HARK-Bird, and normalize these values so that their range is from 0 to 1.

2. We calculate the sum of the CVR values corresponding to the frequency range from 2.6 to 3.1 kHz, which is further divided by the sum of the entire values. We call this value the relative CVR (RCVR).

3. Each sound source is classified as a vocalization of cicada (or bird) if its RCVR is above (or below or equal to) the threshold value 0.2. The sound is regarded as noise if its RCVR is less than a small threshold (=0.0 in this case).

Note that used the normalized value in order to exclude the misclassification of cicadas due to other insect noises, and we adopted this threshold value because there were two peaks on both sides of the threshold in the frequency distribution of RCVR. While it is inevitable that this automatic but rough procedure can lead to misclassifications, we think that the results are enough to illustrate the basic tendency of their acoustic behaviors.

In order to investigate inter-specific interactions between birds and cicadas, we compared the temporal changes in vocal activities of birds and cicadas. Their activity in each 300-second time segment is calculated as the total duration of localized sounds in the segment, which is normalized by the maximum value overall segments.

3 Preliminary results

3.1 Soundscape analysis

Fig 3 shows (a) acoustic index-based and (b) DOA-based soundscapes. Each panel corresponds to a 13-minute recording. In (a), we can see regions colored with yellow (a mixture of red and blue) in the intermediate frequency range around 2.5-3 kHz. This means that ACI and 1-H[t] reflected similar sound events. These regions indicate bird vocalizations because high values of both indices reflect large temporal changes in the amplitude within short time periods. Actually, we see repetitions of short and high-frequency vocalizations around the corresponding time periods in (b). For example, we see some songs of Blue-and-white Flycatcher (purple), Red-billed leiothrix (orange), and Japanese Bush Warbler (green) in Fig. 4 (bottom, 14:16-14:19). The vocalizations were colored with similar colors among vocalizations of each species but they tended to be different between species. This implies that a single individual might be singing in a different direction for each species. However, their song colors tended to be biased strongly by simultaneously vocalizing songs of cicadas in other time periods, and thus the method needs further improvement.

We also see in (a) that there exist blue (CVR) narrow regions around 3 kHz. They reflect songs of cicadas as expected, and the corresponding clusters of songs were indicated with quite different colors in (b). This means that multiple individuals of cicada were expected to be singing in different directions alternately in this recording as illustrated in an example situation in Fig 4 (top, 13:49-13:52).

3.2 Vocal activity analysis

Fig. 5 shows the distribution of vocalizations in the space of time and direction of arrival, which were classified into bird and cicada vocalizations. The red and blue bars represent vocalizations of birds and cicadas, respectively. We found that multiple individuals of both birds and cicadas were vocalizing during the recording since their vocalizations were localized in various directions. We also see that there were time durations tended to be dominated by cicadas (e.g.

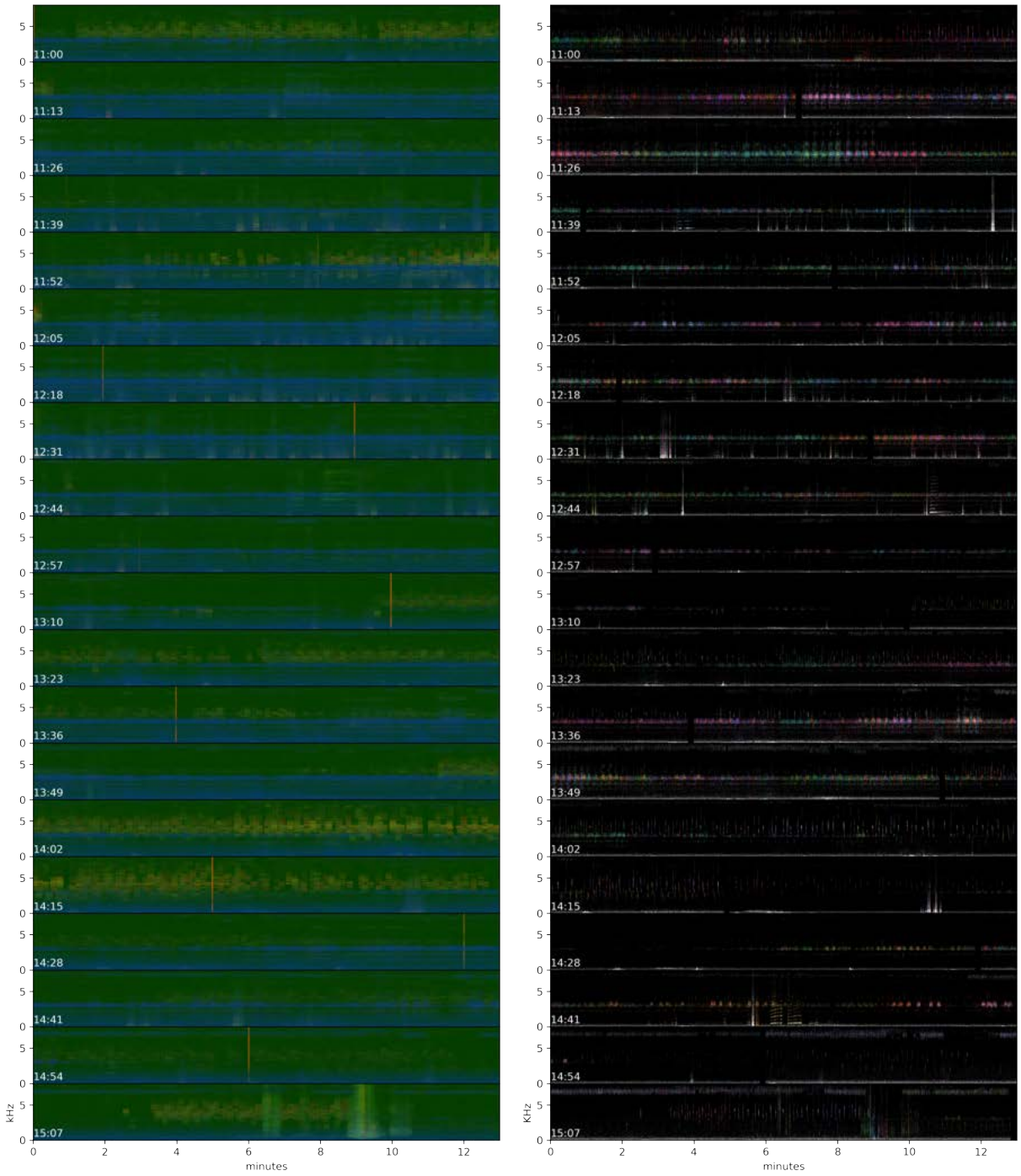

Figure 3: The acoustic index-based (left) and DOA-based (right) spectrograms for an about 4-hour recording.

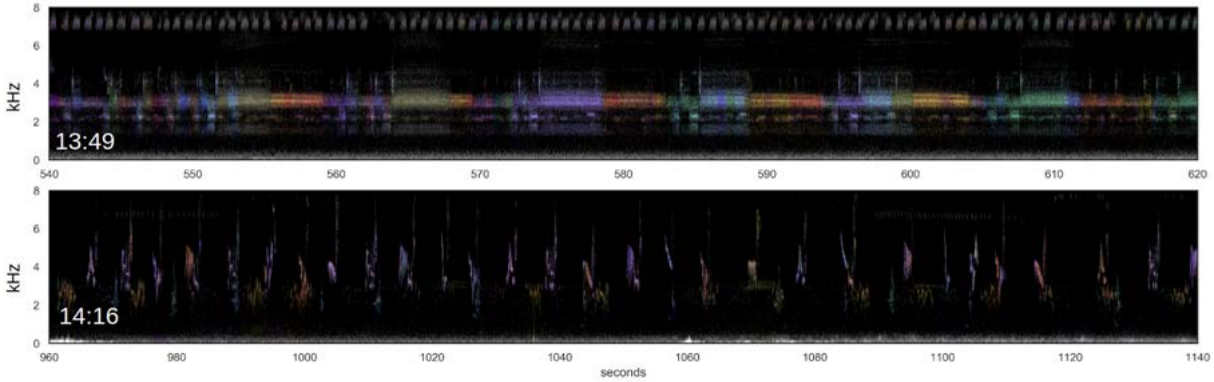

Figure 4: Examples of DOA-based spectrograms showing songs of birds (bottom) and cicadas (top).

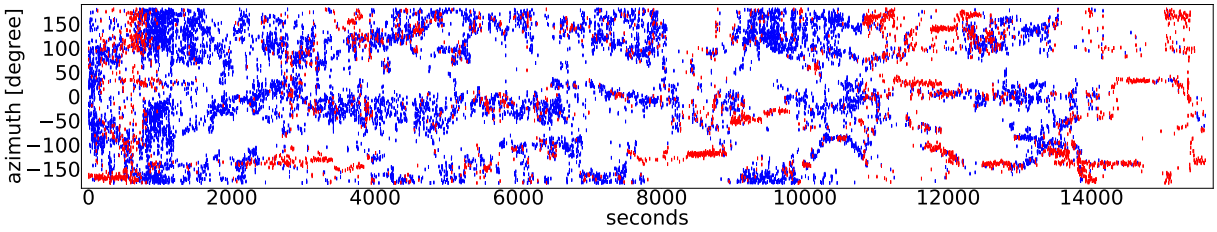

Figure 5: The directional distribution of localized vocalizations of birds (red) and cicadas (blue).

1000-2000 seconds) and one tended to be dominated by birds (e.g. 11000-12000 seconds).

Fig. 6 shows the changes in the vocal activity of birds and cicadas defined in Section 2.4. In the first half of the recording, the cicadas vocalized actively but the birds were relatively quiet, except for the first 900 seconds. On the other hand, the birds vocalized actively and the cicadas were gradually getting quiet in the latter half of the recording. In addition, it is suggested that there were vocal turn-taking between birds and cicadas at intervals of 5 to 15 minutes in that their activities repeated increased alternately. This could be an overlap avoidance of vocalizations between them because the frequency bands of vocalizations uttered by the birds and cicadas in this recording were relatively close. However, we need detailed analyses based on more sophisticated vocalization classification procedures.

We also observed intra-specific turn-takings between cicada individuals. Figure 7 shows an example of turn-taking situation. In this duration, multiple cicadas vocalized in some directions. The cicadas vocalized at -50 degrees and -100 degrees alternately in the first half. The cicada vocalized at -100 degrees and the positive directions (100 and 150) alternately. Both of them imply the occurrences of turn-takings among multiple individuals.

4 Conclusion

This paper discussed an application of robot audition techniques to a soundscape analysis of a complex situation of vocalizing birds and cicadas in early summer. We showed that two types of false-color spectrograms based on acoustic indices and direction of arrival can illustrate the overall dynamics of their acoustic behaviors. While the methods still need improvement, the preliminary quantitative analysis of their vocal activities implied that there might exist temporal avoidance behaviors among these birds and cicadas. We also found that there might also exist intra-specific turn-takings between cicada individuals.

Acknowledgements

We thank Naoki Takabe (Nagoya University) for supporting field recordings. This work was supported in part by JSPS/MEXT KAKENHI: 20H00475, 19KK0260, JP18K11467, and JP17H06383 in #4903 (Evolinguistics).

References

[1] A. Farina and S. H. Gage. *Ecoacoustics: The Ecological Role of Sounds*. John Wiley and Sons, 2017.

[2] R. Suzuki, Sumitani S. Zhao, H., S. Matsubayashi, K. Arita, T. Nakadai, and H. G. Okuno. Visualizing directional soundscapes of bird vocalizations using robot audition techniques. In *Proceedings of Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII 2021)*, in press.

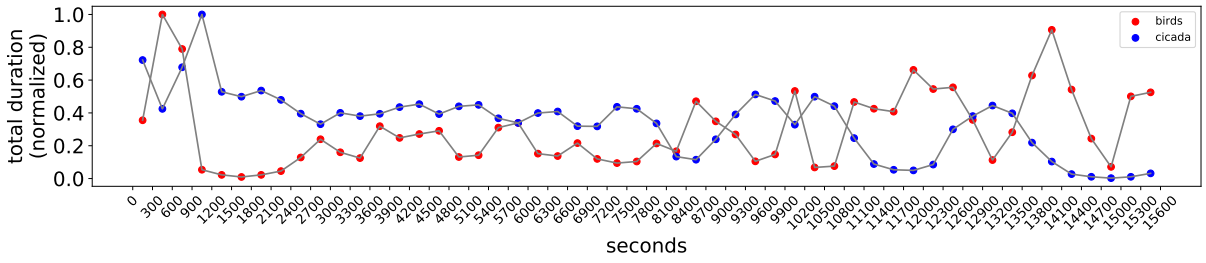

Figure 6: The changes in vocalization activities of birds (red) and cicadas (blue).

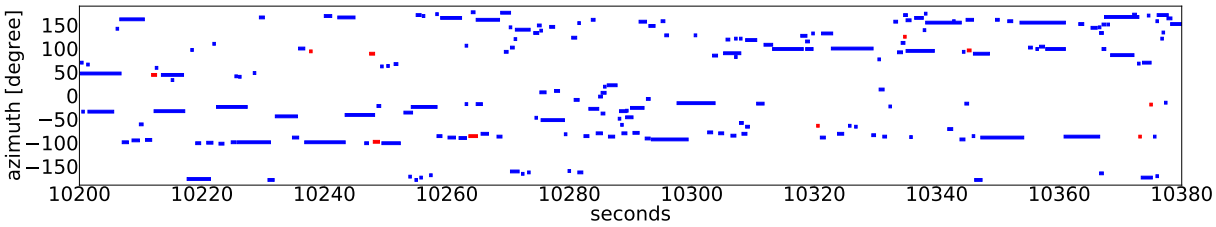

Figure 7: An example of turn-taking process of multiple cicada individuals.

[3] Shinji Sumitani, Reiji Suzuki, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai, and Hiroshi G. Okuno. Fine-scale observations of spatio-spectro-temporal dynamics of bird vocalizations using robot audition techniques. *Remote Sensing in Ecology and Conservation*, rse2.152, 2020.

[4] M. Towsey, L. Zhang, M. Cottman-Fields, J. Wimmer, J. Zhang, and P. Roe. Visualization of long-duration acoustic recordings of the environment. *Procedia Computer Science*, 29:703–712, 2014.

[5] M. Towsey, E. Znidersic, J. Broken-Brow, K. Indraswari, D. M. Watson, Y. Phillips, A. Truskinger, and P. Roe. Long-duration, false-colour spectrograms for detecting species in large audio data-sets. *Journal of Ecoacoustics*, 2:IUSWUI, 2018.

[6] M. L. Cody and J. H. Brown. Song asynchrony in neighbouring bird species. *Nature*, 222:778–780, 1969.

[7] H. Brumm. Signalling through acoustic windows: nightingales avoid interspecific competition by short-term adjustment of song timing. *Journal of Comparative Physiology A: Neuroethology*, 192:1279–1285, 2006.

[8] P. J. Hart, R. Hall, W. Ray, A. Beck, and J. Zook. Cicadas impact bird communication in a noisy tropical rainforest. *Behavioral Ecology*, 26:839–842, 2015.

[9] H. Slabbekoorn and M. Peet. Birds sing at a higher pitch in urban noise. *Nature*, 424, 2003.

[10] K. Nakadai, H. G. Okuno, and T. Mizumoto. Development, Deployment and Applications of Robot Audition Open Source Software HARK. *Journal of Robotics and Mechatronics*, 27:16–25, 2017.

[11] R. Schmidt. Bayesian nonparametrics for microphone array processing. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.

[12] S. Sumitani, R. Suzuki, S. Matsubayashi, K. Arita, T. Nakadai, and H. G. Okuno. An integrated framework for field recording, localization, classification and annotation of birdsongs using robot audition techniques - harkbird 2.0. In *Proceedings of ICASSP 2019*, pages 8246–8250, 2019.

[13] N. Pieretti, A. Farina, and D. Morri. A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators*, 11:868–873, 2011.