人工知能学会第二種研究会資料
AIチャレンジ研究会
SIG-Challenge-061-02

A soundscape analysis of bird and cicada vocalizations based on azimuth and elevation localization using robot audition techniques

Hao Zhao[1*] Reiji Suzuki[1] Takaya Arita[1] Kazuhiro Nakadai[2]

[1] Nagoya University
[2] Tokyo Institute of Technology

Abstract: This study aims to apply robot audition techniques to investigate the natural soundscape of forest animal vocalizations based on azimuth-elevation estimation of sounds. We focus on two recordings in an experimental forest in Japan, where a 16-channel semi-spherical microphone array had been placed, and birds and cicadas dominated the soundscape. We manually annotated the localization results obtained from sound source localization and separation based on HARK, using HARKBird. Then we conducted a preliminary analysis on their vocal activity and the azimuth and elevation information. The result showed the difference in the soundscape patterns of their vocalizations in the azimuth-elevation space: a few birds tended to stay and sing at their song posts, and cicadas formed complex and bursty singing behavior, changing their positions. The behavior of cicada vocalizations may have reduced the vocal activity of birds that share the frequency ranges of their vocalizations while needs further detailed analyses to conclude.

Ecoacoustics [1] is an interdisciplinary field that investigates natural and human sounds and their relationships with an environment, which contributes to long-term ecosystem monitoring, habitat conservation, biodiversity assessment, and ecosystem management.

When considering the roles of sounds in ecoacoustics, the soundscape is an important concept that refers to the combination of sounds that arise from both natural and artificial environments [2]. Extracting a precise spatio-temporal structure of a soundscape and grasping the soundscape dynamics are essential for ecoacoustics to track active interactions among individuals and grasp the overall properties of the soundscape. However, it is not straightforward to grasp such a complex acoustic structure with a standard autonomous recording unit because it is hard to extract the spatio-temporal information of multiple sounds occurring in natural environments. Thus, there is increasing interest in microphone arrays to localize animal vocalizations [3, 4].

We have proposed and discussed novel applications of robot audition techniques to investigate the 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 the robot audition software HARK (explained later) [5, 6]. We visualized and quantified the directional and 2D-spatial patterns of bird vocalizations in various contexts such as response behaviors in playback experiments [7, 8], for example.

As a next approach, we have been focusing on the soundscape dynamics of multiple classes of species dominating the soundscape of forests in early summer: birds and cicadas. 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 [9, 10, 11], other animals[12] and abiotic noise [13]. 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 [12]. We used a method to classify their vocalizations using three ecoacoustic indices (acoustic complexity index, temporal entropy, and acoustic cover), then illustrated their temporal vocal activities, measured as the total song duration in each time segment [14]. As a proof-of-concept, we applied this to three scenarios of recordings to see if there exist inter-specific interactions between birds and cicadas and replayed the vocalizations of the cicadas to observe the effect on their vocalization activities. The preliminary analysis implied that there might exist temporal overlap avoidance behaviors between birds and cicadas, and replayed songs of cicadas may reduce the activity of birds.

While the above previous work is based on the azimuth estimation of acoustic events, there also exist variations in elevations among natural sound sources. Pekin et al. demonstrated the use of LIDAR-derived metrics and sound recordings for identifying canopy structural attributes supporting high acoustic diversity in a neotropical forest environment and showed that the composition of acoustic frequency bands and acoustic diversity are strongly linked with the vertical structure of the local forested environments [15]. Zezhou et al. used the soundscape mapping to explore the habitat selection of bird communities in the context of spatial-temporal structural changes and showed the urban forest vertical structure had a great ef-

*Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601
E-mail:zhao.hao.y0@s.mail.nagoya-u.ac.jp

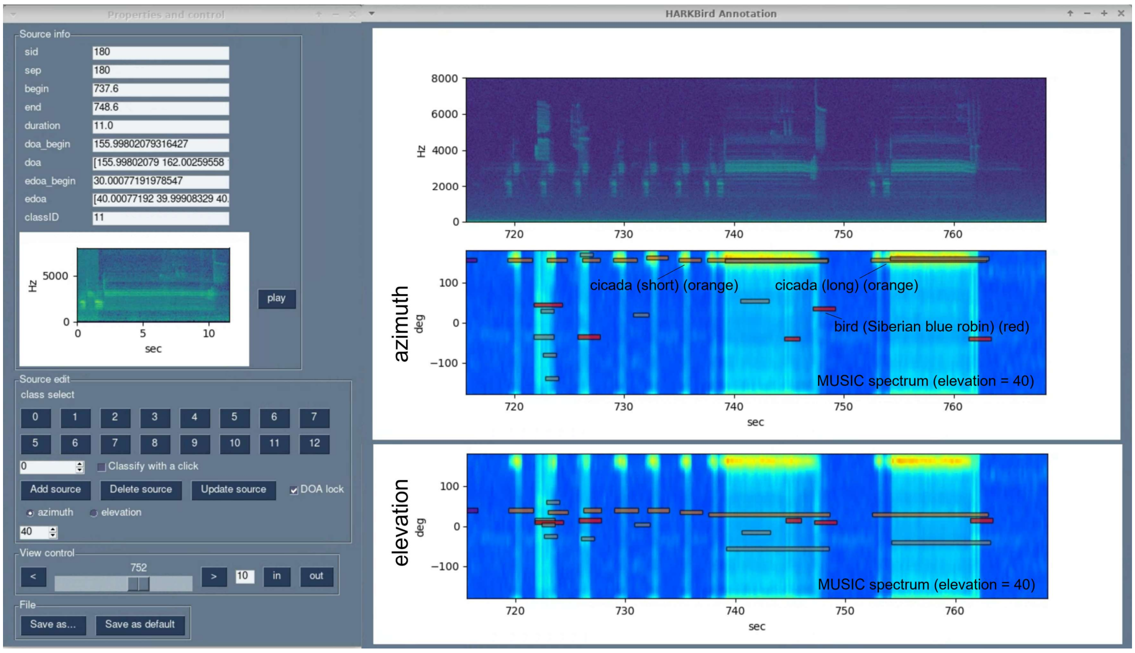

Figure 1: A snapshot of the annotation tool of HARKBird and examples of localized bird and cicada vocalizations.

fect on bird activities [16]. Yamamoto et al. visualized the spatio-temporal structure of soundscape composed of antholophony, biophony and geophony using the azimuth-elevation estimation of sound sources and their separation based on HARK, and their classification based on a recurrent convolutional neural network trained for bird vocalization classification [17]. They discussed that these three classes of sounds can be classified by their elevation information. These studies clarified the importance of elevation information for field observation of soundscape dynamics.

This study aims to further consider the application of robot audition techniques to the observation of the natural soundscape of forest animal vocalizations based on azimuth-elevation estimation of sounds. We focus on the recordings in an experimental forest, Nagoya University, Japan, on June 2021, where a 16-channel semi-spherical microphone array (Chirpy type-S; System in Frontier Inc.) had been placed, and birds and cicadas dominated the soundscape. This waterproof and standalone microphone array enables us to record and localize the out-door soundscape in azimuth and elevation angles. As a preliminary approach, we manually annotated the localization results of two recordings obtained from sound source localization and separation based on HARK, using HARKBird. Then, we conducted a preliminary analysis on their vocal activity and the azimuth and elevation information of their vocalizations. The analysis implied that there might exist temporal overlap avoidance behaviors between birds and cicadas, as expected from our previous results, constructing the distinctive spatial structure of vocalization dynamics in the azimuth-elevation space of the forest soundscape.

1 Materials and methods

1.1 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 [18]. See the website of HARK for detail[1].

We used HARKBird [5, 6], a collection of Python scripts for bird song localization, to estimate the DOA of sound sources in recordings, using sound source localization and separation functions in HARK. We adopted the current version (3.0b), based on PySimpleGUI, which enables us to visualize and annotate the sound source distribution in the azimuth-elevation space as shown in Fig. 1.

The employed sound source localization algorithm is based on the multiple signal classification (MUSIC) [19] using multiple spectrograms obtained by short-time Fourier transformation (STFT). The MUSIC method is a widely used high-resolution algorithm and is based on the eigenvalue decomposition of the correlation matrix of multiple signals from a microphone array. We adopted the standard eigenvalue decomposition (SEVD) MUSIC method implemented as one of the sound source localization methods in HARK. All localized sounds are separated to the sounds as wave files (16 bit, 16 kHz) using geometric high-order decorrelation-based source separation (GHDSS) method [20], which is also implemented

[1] https://hark.jp

in HARK. For more details on HARKBird[2], see [5, 6]. To optimize localization performance, we can adjust some parameters of HARKBird, such as the source tracking and the lower bound frequency for MUSIC, to reduce noise, etc.

Figure 2: An experimental field and a recording node (Chirpy type-S).

1.2 Field recording

On June 2021, we conducted the recording experiments to observe the vocalizations of cicada and bird individuals at our field site in the Inabu field, the experimental forest of Field Science Center, Graduate School of Bioagricultural Sciences, Nagoya University, in central Japan (Figure 2). The forest is mainly composed of conifer plantations (Japanese cedar, Japanese cypress, and red pine), with small patches of broadleaf trees (quercus, acer, carpinus, etc.).

During the recording, the common bird and cicada species were known to vocalize during early summer. The Siberian blue robin (*Larvivora cyane*) was the species that mainly dominated the soundscape. There were also a single species of cicadas (*Terpnosia nigricosta*). The vocalization of this species has a unique structure that is composed of repetitions of introductory short components like frog calls, and subsequent main song components like songs of poplar evening cicadas. These bird and cicada species share the frequency ranges of their vocalizations.

We used a 16-channel microphone array (Chirpy type-S; System in Frontier Inc.) system for out-door long-term recording. The system is composed of a hemispherical microphone unit connected to the control box. The water-proofed 16 microphone channels were placed in a spiral manner on the surface of the unit, which allows us to estimate both azimuth and elevation angles of the localized sounds, using HARKBird. It can also record the posture information of the unit as an additional channel in the wave file. The control box has functions of power supplies based on multiple USB batteries, scheduled recording, and the Wi-Fi-based server system that allows us to configure the schedule and conduct manual recordings via a client software for Windows PCs.

We conducted experimental recordings on consecutive days this month. In this paper, we picked two recording sessions for analysis as examples of the natural soundscape. The first 10-min recording started at 9:00 am, on June 26th. The other 20-min recording started at 9:20 am, on June 24th in which sometimes the experimenter's activity sounds were included.

1.3 Sound source localization, separation, and manual annotation

We used the HARKBird to export the information on localized sound sources (i.e., the beginning and end time, DOA (azimuth and elevation), and its separated sound file (wave file)). To extract the overall spatio-temporal structure of the soundscape, we adjusted the other parameters in HARKBird to localize these vocalizations that allow us to localize most sound sources around the microphone array.

After sound localization and separation, we manually annotated the vocalization events of cicadas and birds by conducting an auditory and visual inspection of the localization and separation results of sound sources. Among the separated sounds, we picked up sources that could be associated with visually recognized bird and cicada vocalizations in the spectrogram. We removed other sources that had no clear associations, which were expected to appear due to irregular noises or effects of other sounds that increased the overall power of MUSIC spectrum. We mainly used the visualized MUSIC spectrum for this purpose. However, there is a possibility that some artifacts were misclassified as cicada vocalizations in this preliminary analyses. Also, there exist a possibility of effects of experimenters' effects on their behaviors.

1.4 Analysis

We visualized the temporal dynamics of the localized vocalizations of birds and cicadas in the azimuth and elevation space, and a circular histogram of the azimuth-elevation angles of localized sounds. These graphs enable us to grasp the overall distribution of their vocalizations.

In order to observe the relationship of the activities between birds and cicadas, we quantified the temporal changes in the vocal activities of birds and cicadas to observe inter-specific interactions between birds and cicadas. Their activity in each 30-second (for the 10-min recording) or 50-second (for the 20-min recording) was calculated as the total duration of localized sounds in the segment.

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

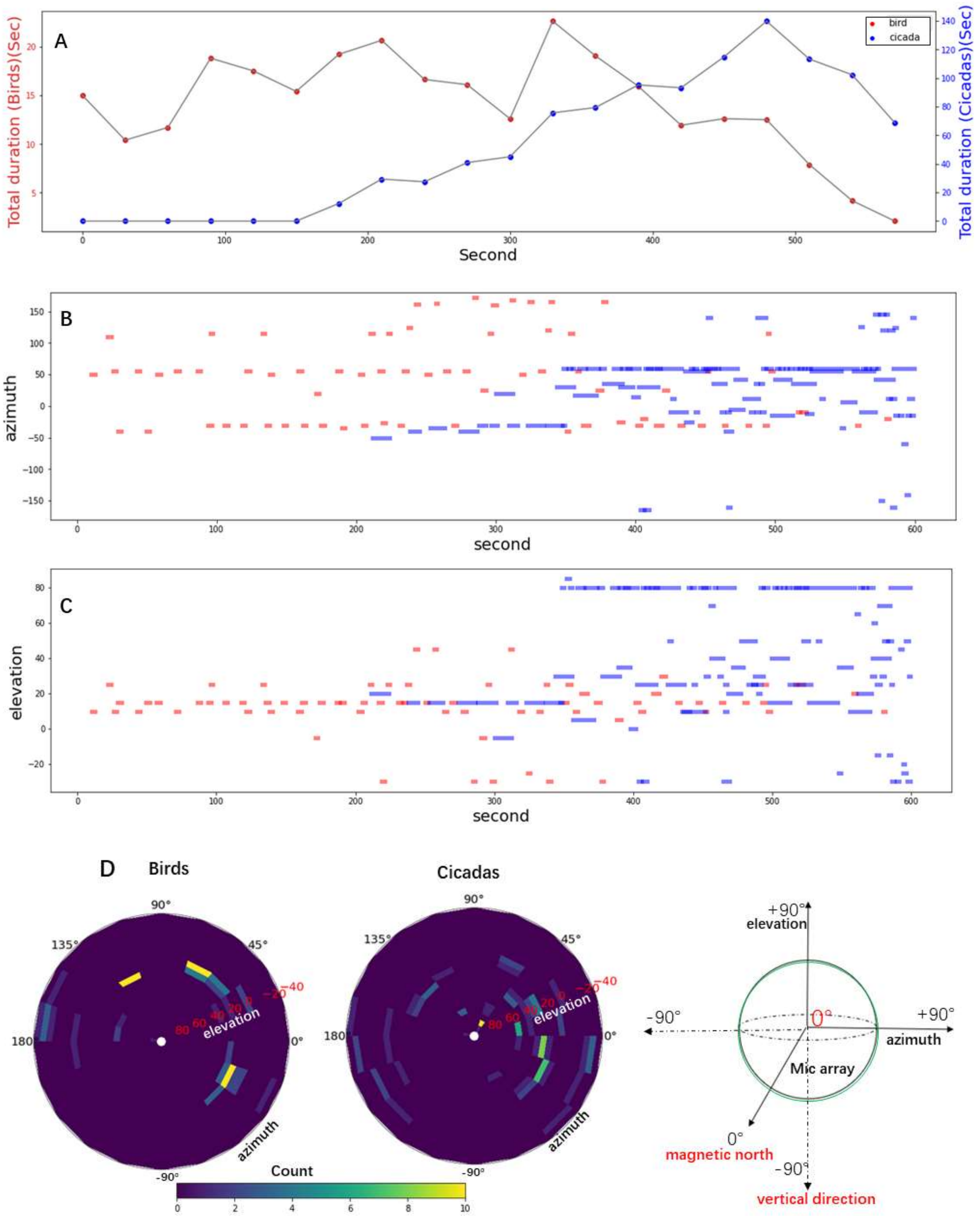

Figure 3: Vocal analysis for 10 minutes (9:00 am-9:10 am, June 26th).

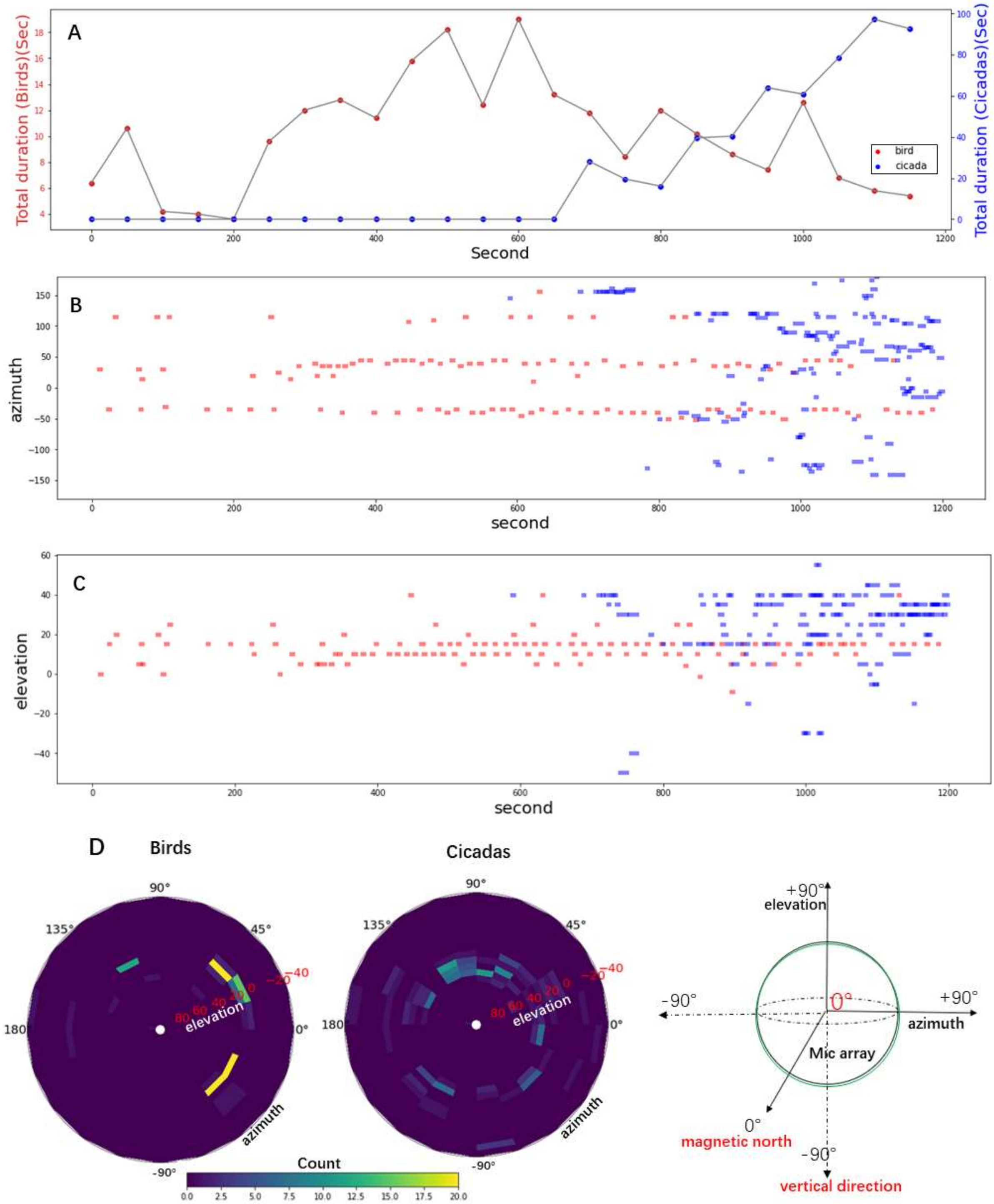

Figure 4: Vocal analysis for 20 minutes (9:20 am-9:40 am, June 24th).

2 Results

Fig. 3 and Fig. 4 show (A) the temporal changes in the vocal activity of the birds (red) and cicadas (blue), (B) the distribution of vocalizations in the space of time and azimuth, (C) the distribution of vocalizations in the space of time and elevation, and (D) the polar histograms of azimuth and elevation angles of bird and cicada vocalizations. All the bird songs were of Siberian blue robins in the two recordings.

In Fig. 3A and Fig. 4A, we can find the birds vocalized actively but the cicadas vocalized relatively quietly in the first half of the recordings. On the other hand, in the later half of both recordings, cicadas vocalized actively but the birds were gradually getting quiet. This indicates that the birds might tend to avoid vocalizing with cicadas.

In B and C of both Fig. 3 and 4, we see a few birds sang repeatedly at consistent positions (e.g., 0-200 seconds in Fig. 3 and 0-700 seconds in Fig. 4), which might be their song posts. Then, one cicada began to sing and the number of singing cicadas gradually increased, forming their chorus or burst. Their songs were composed of short and long components and their singing positions appeared to change both in azimuth and elevation. When their vocalization activity increased, some birds stopped singing or sang less frequently (e.g., after 400 seconds in Fig. 3 and after 900 seconds in Fig. 4), which made the activity of birds relatively small. This supports that these birds might have avoided singing with cicadas. However, further analyses based on long-term data are necessary to conclude.

Fig. 3D and 4D well illustrated the overall soundscape structures of bird and cicada vocalizations in the space of azimuth-elevation angles. We see that the birds at different azimuths vocalized at similar elevation angles. It reflects, in these particular cases, birds tended to sing in the middle of the trees because this is the typical behavior of Siberian blue robins. Another interesting fact is that their soundscape structures were similar between these two different days, which indicates that their habit use were consistent. On the contrary, we see that the soundscape structures of cicada vocalizations were more complex in the sense that there were large variations and frequent changes of localized positions in both azimuth and elevation angles. Some cicada sounds even came from places lower than the microphone because the microphone was placed around a small ridge. Also, the structures varied between both recordings, implying that their temporal dynamics were also complex. We expect that longer-term analyses will enable us to discuss the relationship between their soundscape structures.

3 Conclusion

This paper applied robot audition techniques to investigate the soundscape dynamics of cicada and bird vocalizations. We focus on two recordings and manually annotated the results of azimuth and elevation localization. Then we conducted a preliminary analysis on their vocal activity and the azimuth and elevation information of their vocalizations. The result showed the difference in the soundscape patterns of their vocalizations in the azimuth-elevation space: a few birds tended to stay and sing at their song posts, and cicadas formed complex and bursty singing behavior, changing their positions. The behavior of cicada vocalizations may have reduced the vocal activity of birds that share the frequency ranges of their vocalizations, which supports previous reports [12]. However, there exists a possibility of mislocalization of cicada vocalizations due to the interference of their loud and multiple vocalizations, which is the future work to be improved. The further analyses of their soundscape based on long-term data also are necessary.

Acknowledgements

We thank Naoki Takabe (Nagoya University) for supporting field experiments. This work was supported in part by JSPS/MEXT KAKENHI: JP21K12058, JP20H 00475, JP19KK0260.

References

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

[2] J.D. Douglas and J.F. Mastin. Soundscape. *Journal of Architectural and Planning Research*, pages 169–186, 1985.

[3] D. Blumstein, D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, K. A. Cortopassi, S. F. Hanser, B. McCowan, A. M. Ali, and A. N. G. Kirshel. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48:758–767, 2011.

[4] T. A. Rhinehart, L. M. Chronister, T. Devlin, and J. Kitzes. Acoustic localization of terrestrial wildlife: Current practices and future opportunities. *Ecology and Evolution*, 10(13):6794–6818, 2020.

[5] R. Suzuki, S. Matsubayashi, K. Nakadai, and H. G. Okuno. HARKBird: Exploring acoustic interactions in bird communities using a microphone array. *Journal of Robotics and Mechatronics*, 27:213–223, 2017.

[6] 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.

[7] Reiji Suzuki, Shinji Sumitani, Naren , Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai, and Hiroshi G. Okuno. Field observations of ecoacoustic dynamics of a japanese bush warbler using an open-source software for robot audition hark. *Journal of Ecoacoustics*, 2:EYAJ46, 2018.

[8] S. Sumitani, R. Suzuki, S. Matsubayashi, T. Arita, K. Nakadai, and H.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.

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

[10] 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.

[11] P. J. Hart, T. Ibanez, K. Paxton, G. Tredinnick, E. Sebastián-González, and A. Tanimoto-Johnson. Timing is everything: Acoustic niche partitioning in two tropical wet forest bird communities. *Frontiers in Ecology and Evolution*, 9:753363, 2020.

[12] 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.

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

[14] H. Zhao, R. Suzuki, S. Sumitani, S. Matsubayashi, T. Arita, K. Nakadai, and H. G. Okuno. Visualizing soundscapes and quantifying interspecific interactions in forest animal vocalizations using robot audition technology. In *Proceedings of 84th Annual Meeting of IPSJ*, volume 2, pages 475–476, 2022.

[15] B. K. Pekin, J. Jung, L. J. Villanueva-Rivera, B. C. Pijanowski, and J.A. Ahumada. Modeling acoustic diversity using soundscape recordings and lidar-derived metrics of vertical forest structure in a neotropical rainforest. *Landscape ecology*, 27(10):1513–1522, 2012.

[16] Z. Hao, C. Wang, Z. Sun, C. K. Vandenbosch, D. Zhao, B. Sun, X. Xu, Q. Bian, Z. Bai, K. Wei, et al. Soundscape mapping for spatial-temporal estimate on bird activities in urban forests. *Urban Forestry & Urban Greening*, 57:126822, 2021.

[17] R. Yamamoto, K. Nishida, K. Itoyama, S. Matsubayashi, and K. Suzuki, R. Nakadai. Soundscape analysis using robot audition open source software hark. In *Proceedings of the Annual Meeting of Ornithological Society of Japan*, 2022.

[18] 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.

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

[20] H. Nakajima, K. Nakadai, Y. Hasegawa, and H. Tsujino. Blind source separation with parameter-free adaptive step-size method for robot audition. *IEEE Transactions on Audio, Speech, and Language Processing*, 18:1476–1485, 2010.