applied sciences

MDPI

Article

Estimating the Soundscape Structure and Dynamics of Forest Bird Vocalizations in an Azimuth-Elevation Space Using a Microphone Array

Reiji Suzuki [1,*], Koichiro Hayashi [2], Hideki Osaka [3], Shiho Matsubayashi [4], Takaya Arita [1], Kazuhiro Nakadai [5] and Hiroshi G. Okuno [6,7]

1 Graduate School of Informatics, Nagoya University, Nagoya 464-8601, Japan
2 School of Informatics and Sciences, Nagoya University, Nagoya 464-8601, Japan
3 toriR Lab., Echizen 915-0242, Japan
4 Graduate School of Engineering Science, Osaka University, Toyonaka 560-8531, Japan
5 Department of Systems and Control Engineering, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8552, Japan
6 Graduate School of Inform, Kyoto University, Kyoto 606-8501, Japan
7 Future Robotics Organization, Waseda University, Shinjuku, Tokyo 169-0072, Japan
* Correspondence: reiji@nagoya-u.jp; Tel.: +81-52-789-4258

Abstract: Songbirds are one of the study targets for both bioacoustic and ecoacoustic research. In this paper, we discuss the applicability of robot audition techniques to understand the dynamics of forest bird vocalizations in a soundscape measured in azimuth and elevation angles with a single 16-channel microphone array, using HARK and HARKBird. First, we evaluated the accuracy in estimating the azimuth and elevation angles of bird vocalizations replayed from a loudspeaker on a tree, 6.55 m above the height of the array, from different horizontal distances in a forest. The results showed that the localization error of azimuth and elevation angle was equal to or less than 5 degrees and 15 degrees, respectively, in most of cases when the horizontal distance from the array was equal to or less than 35 m. We then conducted a field observation of vocalizations to monitor birds in a forest. The results showed that the system can successfully detect how birds use the soundscape horizontally and vertically. This can contribute to bioacoustic and ecoacoustic research, including behavioral observations and study of biodiversity.

Keywords: bird song; soundscape; ecoacoustics; sound source localization; robot audition; HARK

Citation: Suzuki, R.; Hayashi, K.; Osaka, H.; Matsubayashi, S.; Arita, T.; Nakadai, K.; Okuno, H.G. Estimating the Soundscape Structure and Dynamics of Forest Bird Vocalizations in an Azimuth-Elevation Space Using a Microphone Array. *Appl. Sci.* **2023**, *13*, 3607. https://doi.org/10.3390/app13063607

Academic Editors: Luis Gracia and Carlos Perez-Vidal

Received: 26 December 2022
Revised: 27 February 2023
Accepted: 4 March 2023
Published: 11 March 2023

1. Introduction

Songbirds are one of the study targets for the purpose of ecoacoustic research [1,2]: an interdisciplinary science that investigates natural and anthropogenic sounds and their relationship to the environment across multiple scales of time and space [3], as well as bioacoustic research. This is because their vocalizations (1) can tell us a suite of useful information about the environment for monitoring, (2) have rich and complex variety of structures [4], which are used as benchmark problems for classification tasks (e.g., BirdCLEF [5]), and (3) enable bird individuals to interact in complex ways, behaving as complex systems [6].

There are several approaches for using microphone arrays to localize bird vocalizations. Rhinehart et al. recently surveyed applications of acoustic localization using autonomous recording units in terrestrial environments [7], and pointed out that ecologists will make better use of acoustic localization; it can collect large-scale animal position data with minimizing the influence on the environment if recording hardware and automated localization and classification software are more available, and their algorithms are improved for outdoor measurement.

Some of these studies focused on both azimuth and elevation estimation [8–10]. Hedley et al. developed a 4-channel microphone array unit by combining two stereo microphones and evaluated the accuracy to estimate azimuth and elevation angles of replayed songs of a few species [9]. The results showed that most of sounds were estimated within 12 degrees of the true direction of arrivals (DOA) in the azimuth angle and 9 degrees in the elevation angle within a range of at least 30 m. It was also discussed that the DOA estimation may improve the ability to assess abundance in biodiversity surveys. However, the experiment was conducted in an open space, and the elevation angle was limited to −10 to 15 degrees from horizontal. As a different approach, Gayk et al. constructed a microphone array system to estimate 3D position of flying songbirds with a wireless microphone array. The system was consisted of four 7-m poles arranged in a 25 m square, and each pole had two microphone channels that are placed on top and bottom of the pole. They adopted a triangulation method based on time-of-arrival differences of a sound recorded at these microphones to cross-correlate and estimate sound position. They showed that both broadcasted bird calls and calls of natural migratory birds were successfully triangulated with the accuracy and estimated accuracy of less than 3 m. In addition, there is increasing interest and development for sound event localization and detection (SELD) of various environmental sounds using microphone arrays and ambisonic microphones [11]. We expect that practical experimental analyses of natural sounds such as bird vocalizations in forests can further contribute to better use of such microphone array-based techniques in natural fields.

We have been proposing that robot audition techniques [12], especially an open source software for robot audition HARK (Honda Research Institute Japan Audition for Robots with Kyoto University), can contribute to bioacoustics and ecoacoustics. It not only provides the DOA estimation of sounds, but also allows us to separate them and perform further signal processing on them, even in real time. We developed HARKBird, a collection of Python scripts for localizing bird songs in fields using HARK [13]. Previously, we confirmed the effects of playback of conspecific song on song or call responses by measuring the changes in their localized direction [14] and changes in their 2D position using a set of microphone arrays [15].

It is recognized that data characterizing the vertical structure of vegetation are becoming increasingly useful for biodiversity applications as remote sensing techniques such as radar and lidar become more readily available [16]. We believe that direct observation of the vertical and horizontal soundscape of vocalizations among birds would also contribute to this field, as well as to bioacoustic analysis of bird behavior. There is initial work on 3D localization of bird songs using multiple microphone array units [10] and observation of nocturnal birds with a single microphone array unit based on the azimuth-elevation estimation [17]. However, we still need to investigate how HARK or HARKBird can estimate both azimuth and elevation angles of bird songs to capture the dynamics and structures of the soundscape of bird songs. In particular, a systematic evaluation of the localization accuracy of elevation angles and an estimation of the structure of soundscape in a realistic situation where multiple bird species are vocalizing are important for the practical use of the system.

This paper aims to demonstrate a systematic evaluation of the localization accuracy of azimuth-elevation angles of replayed bird vocalizations in a practical forest environment, and show an example field observation of the structure and dynamics of birdsong soundscape. For this purpose, we use a self-developed 16-channel microphone array, called DACHO, using HARK and HARKBird. Suzuki et al. [18] used the same microphone array to conduct spatiotemporal analysis of acoustic interactions between great reed warblers (*Acrocephalus arundinaceus*). They conducted a 2D localization of their vocalizations using multiple arrays and estimated the location of two individuals' song posts with mean error distance of 5.5 ± 4.5 m from the location of observed song posts. They then evaluated the temporal localization accuracy of the songs by comparing the duration of localized songs around the song posts with those annotated by human observers, with an accuracy score of average 0.89 for one bird that stayed at one song post. However, the localization accuracy

of songs in the elevation angle was not evaluated, and thus a systematic analysis of the accuracy of elevation angle estimation in field conditions would supplement and strengthen our knowledge about the application of robot audition techniques to ecoacoustic research.

We used a single microphone array unit because it is a minimal system and its cost is low for field deployment. We think that sound source localization is useful to passively monitor auditory behaviors of rare or nocturnal birds. Localized results can be used to estimate the abundance and the distribution of those birds. The high portability and low deployment cost are both essential in such a case.

First, we evaluated the accuracy in estimating the azimuth and elevation angles of bird vocalizations replayed from a loudspeaker on a tree, 6.55 m above the height of the array, from different horizontal distances in a forest. The results showed that the localization error of azimuth and elevation angle was equal to or less than 5 degrees and 15 degrees, respectively, in most of cases when the horizontal distance from the array was equal to or less than 35 m. We then conducted field observation of vocalizations to monitor birds in a forest. The results showed that the system can successfully capture how birds use the soundscape horizontally and vertically. This can contribute to bioacoustic and ecoacoustic research, including behavioral observations and study of biodiversity.

The organization of the paper is as follows: We firstly introduce two cases of experimental trials: a speaker test and field observation of soundscape dyamics of bird vocalizations, and introduce the sound source localization method based on HARK and HARKBird in Section 2. Then, we show experimental results of the two trials in Section 3, and finally summarize and discusses the significance of the findings and their implications for further contribution to ecoacoustics and related fields in Section 4.

2. Methods

2.1. Speaker Test

We conducted a speaker test to investigate whether and how bird vocalizations can be localized in a forest environment using azimuth and elevation angles. The experiment was conducted at Nagano park, Kawachinagano, Osaka, Japan on 3 December 2018 (Figure 1). Figure 2 shows a schematic diagram of two experimental setups. In Experiment 1, we placed a loudspeaker on a tripod (height = 1.3 m). A 16-ch microphone array DACHO (WILD-BIRD-SONG-RECORDER; SYSTEM IN FRONTIER Inc., Tokyo, Japan) was also placed on a tripod. The array was specifically developed for bird observations in the field. It consists of 16 microphones, arranged within an egg-shaped frame, which is 17 cm in height and 13 cm in width. It records using a 16-channel, 16 bit, 16 kHz format. Recorded raw data are stored in SD cards and can be exported in wave format for further analysis. One can schedule a recording by preparing the time settings in a micro-SD card. See [18] for more detail and an example of using this microphone array in open fields. We changed the distance between the loudspeaker and the microphone array from 0 to 65 m, with an interval of 5 m, by moving the microphone array along a straight path. This is because the maximum length of the ridge that could be considered straight was 65 m around the loudspeaker. Within this distance, a spacing of 5 m was chosen as it was sufficient to measure the effect of the difference in loudspeaker height and the horizontal difference between the array and the loudspeaker.

We replayed a sound file containing four vocalizations of Scaly Thrush (*Zoothera dauma*) at each location as shown in Figure 3. The distance between the loudspeaker and the microphone was 30 m (Experiment 1). In this figure, four vocalizations of the replayed songs were localized successfully, and at the same time, other sound sources were localized around 1 and 8 s. This species is known to sing this type of songs mainly at night. In this experiment, we adopted this vocalization as the playback sound, to simulate observations of such nocturnal vocalizations, which are not easily observed by other methods such as video recordings.

In Experiment 2, we attached the loudspeaker on a tree, 6.55 m above the height of the microphone array. This is because it was the maximum height at which we could

safely place the loudspeaker and at which we could study the effect of the height of the loudspeaker on the localization accuracy of the replayed sound. We performed the same speaker experiment as in Experiment 1.

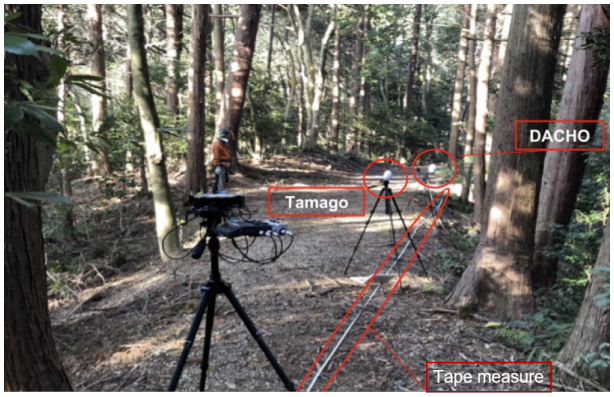

Figure 1. A snapshot of the experiment. We used DACHO, a 16-ch microphone array, for recording replayed songs from a loudspeaker.

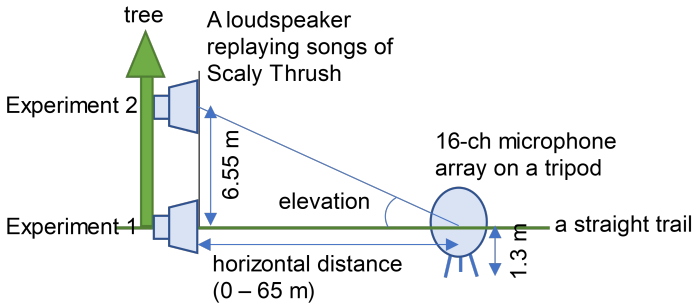

Figure 2. A schematic image of the experimental condition.

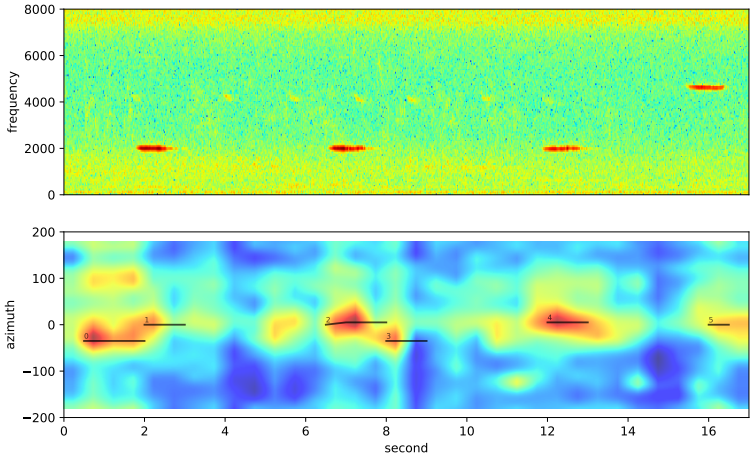

Figure 3. An example of the recording of a replayed sound (**top**) and the localization results (**bottom**). We used the latter part of a replayed sound that include four vocalizations of Scaly Thrush, which is shown in the top figure. The bottom figure shows a heat map of the MUSIC spectrum, whose value represents the strength of sound existence in the corresponding direction. Each black line represents the duration and direction of a localized sound.

2.2. Field Observation of Soundscape Dynamics of Bird Vocalizations

We also conducted a field observation with the same microphone array set up in the Inabu field, the experimental forest of the Field Science Center, Graduate School of Bioagricultural Sciences, Nagoya University, Japan. The forest is mainly a conifer plantation

with small patches of broad leaf trees. We placed a DACHO on a path around a patch of broad leaf trees as shown in Figure 4.

Figure 4. A panorama (**A**) and 360 degree (**B**) photos around the microphone array.

Recording was conducted from 4 April to 7 May 2018. Common bird species actively vocalized during the breeding season here. In particular, Blue-and-white Flycatchers (*Cyanoptila cyanomelana*) tend to sing, advertising their territories, on top of tall trees in their territories. We focused on a 1000-s recording from 8:00 AM on 3 May (Figure 5), where such a typical pattern of bird vocalizations was observed.

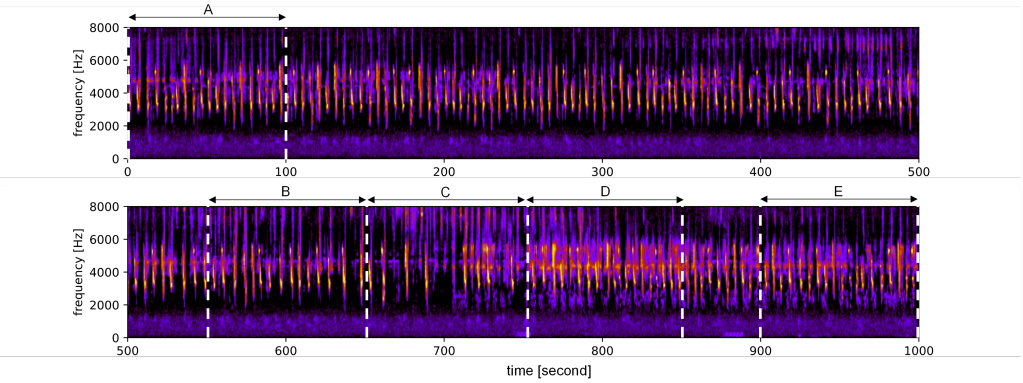

Figure 5. A spectrogram of the whole recording and 5 time slots on which we focus in the analysis.

2.3. Bird Song Localization Using HARKBird

We used HARKBird 2.0 [13], 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.

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 the sounds as wave files (16 bit, 16 kHz) using geometric high-order decorrelation-based source separation (GHDSS) method [20], which is also implemented in HARK. For more details on HARKBird (http://www.alife.cs.i.nagoya-u.ac.jp/~reiji/HARKBird/), accessed on 25 December 2022, see [13,21] and on HARK, see Nakadai et al. [12]. In order 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.

We used a transfer function of the microphone array created from a numerical simulation based on the geometry of the channels of the microphone array using HARKTool5, assuming that there no effects of the body of the array unit on sound transmission. The resolution of DOA for azimuth angle was 5 degrees. The resolutions for elevation angles were 5 and 15 degrees for a speaker test and a bird observation, respectively.

For the DOA estimation of replayed vocalizations in the speaker test, we used the limited frequency range (1800–5000 Hz) for sound source localization, which included the replayed songs. We found that the amplitude of the replayed vocalizations became weaker the farther as the microphone array was from the speaker. Therefore, we gradually decreased the threshold parameter (THRESH), which determines the minimum value of the MUSIC spectrum to detect a sound source, from 29 to 20 with increasing distance. we determined these threshold values empirically according to the acoustic condition around the microphone array. However, this resulted in HARKBird localize vocalizations of other bird species more frequently. Therefore, we manually selected the localized sounds that were detected as replayed vocalizations, and excluded other localized sounds from the analyses. We also lowered the threshold in degree to distinguish multiple sound sources in different directions when there were other sound sources in closer directions to the replayed songs. This is to avoid recognizing them as a part of the replayed vocalizations. We used default values for the other parameters of HARKBird.

For the field observation, we focused on five 100-s time slots (A–E) during which a Blue-and-white Flycatcher (Figure 5) sang on top of tall trees, along with other species such as Varied Tit (*Sittiparus varius*) and Coal Tit (*Periparus ater*). We focused on the behavioral changes in the individual of Blue-and-white Flycatcher, and chose the durations that well illustrated his different behavioral patterns.

We adjusted parameters of HARKBird to localize their songs and exclude other sound sources. We plotted the distribution of songs in the polar-coordinate system representing the azimuth-elevation space for each slot. We then calculated the elevation and azimuth variations of the localized sounds to see if such statistical metrics could reflect the soundscape structures of bird songs.

3. Results

3.1. Speaker Test

Figure 6 shows the estimated azimuth (left) and elevation (right) of the replayed vocalizations in Experiment 1 (top) and 2 (bottom). A red line represents the expected value. Each box plot represents the distribution of localized values when the microphone array was placed x m from the loudspeaker. In Experiment 1, the expected azimuth and elevation angles were 0 degrees. The errors of observed azimuth were equal to or less than 5 degrees when the distance was equal to or less than 50 m. The errors of observed elevation were equal to or less than 10 degrees except in the case of 15 m distance. The slightly larger error when the distance = 15 m was expected to be due to the vocalization of another species (Brown-eared Bulbul (*Hypsipetes amaurotis*)). The large error when the distance = 0 m is expected to be due to that the DOA substantially became different among localized sounds because they can change drastically even with small noise if the microphone is right under the loudspeaker. Thus, both the elevation and azimuth angles of replayed vocalizations were successfully estimated in this experiment.

In Experiment 2, the expected azimuth was 0 degrees, while the expected elevation decreased inversely proportional to the distance of 90 degrees, as shown in Figure 6 (bottom right). The errors of the observed azimuth were equal to or less than 5 degrees until the horizontal distance was equal to or less than 35 m, while it became larger than Experiment 1. This result was expected because the net distance between the microphone and the loudspeaker was larger. This was also expected because other species were vocalizing in the same direction as the speaker, causing the localization of replayed sounds to deviate from the expected value, which was sometimes observed in Experiment 2.

The observed elevation also reflected well the expected value; it decreased inversely proportional to the distance and errors were less than 15 degrees, except in some cases (e.g., 15 or 40 m away). We expect that the errors can be reduced if we adopt a transfer function with a higher resolution of elevation angles.

Overall, we were able to correctly estimate both the elevation and azimuth of bird vocalizations even when the songs were far away from the microphone array, if there were no other vocalizations or sounds in the similar direction as the target sounds.

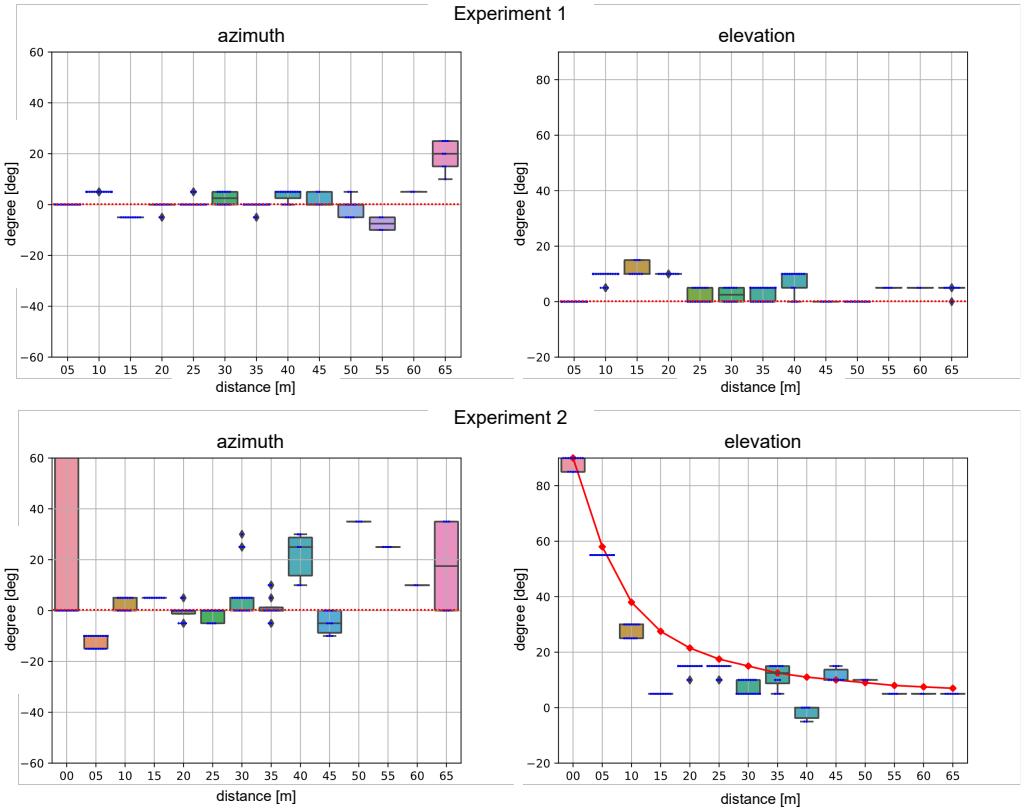

Figure 6. An estimated azimuth (**left**) and elevation (**right**) of replayed songs in Experiment 1 (**top**) and 2 (**bottom**).

3.2. Field Observation of Soundscapes of Bird Vocalizations

Figure 7 shows the song distribution on polar coordinate of azimuth (angle) and elevation (radius) in each time slot (A–E) in Figure 5. Each plot represents the direction of localized bird song (with annotations for species names). We mainly observed songs of Blue-and-white Flycatcher (*Cyanoptila cyanomelana*), Varied Tit (*Sittiparus varius*) and Coal Tit (*Periparus ater*). When we focus on a Blue-and-white Flycatcher, the individual tended to sing at much higher positions than other species, repeatedly moving to other high positions and singing a few times over B to C and returning to the starting position (A) in D. This reflects the fact that this species tends to sing on high trees along streams in his territory.

In contrast, the songs of the other two species tended to be localized at lower elevation angles, suggesting that they tend to sing at lower positions around the microphone. We also see that the localized positions formed multiple clusters, indicating they tended to move slightly. Thus, we could quantitatively observe the spatial structure of the soundscape in which one species tended to occupy the high elevation range, while the other species occupied the lower range.

Table 1 shows some indices on the localized sounds in each time slot. We observe see several changes in the soundscape structure of bird songs. The number of localized songs gradually increased over the time slots, indicating that actively singing individuals (i.e., Varied Tit) entered this acoustic scene. Elevation angle variation became smallest at

C, indicating that the Blue-and-white Flycatcher was probably at a relatively distant tree, considering this species tends to sing on top of a tree. The high values of azimuth variation in B and C reflect that the Blue-and-white Flycatcher moved during the period and other species sang in the opposite direction. Thus, we can grasp the dynamics of the soundscape structures around the microphone array by looking at the changes in these types of indices.

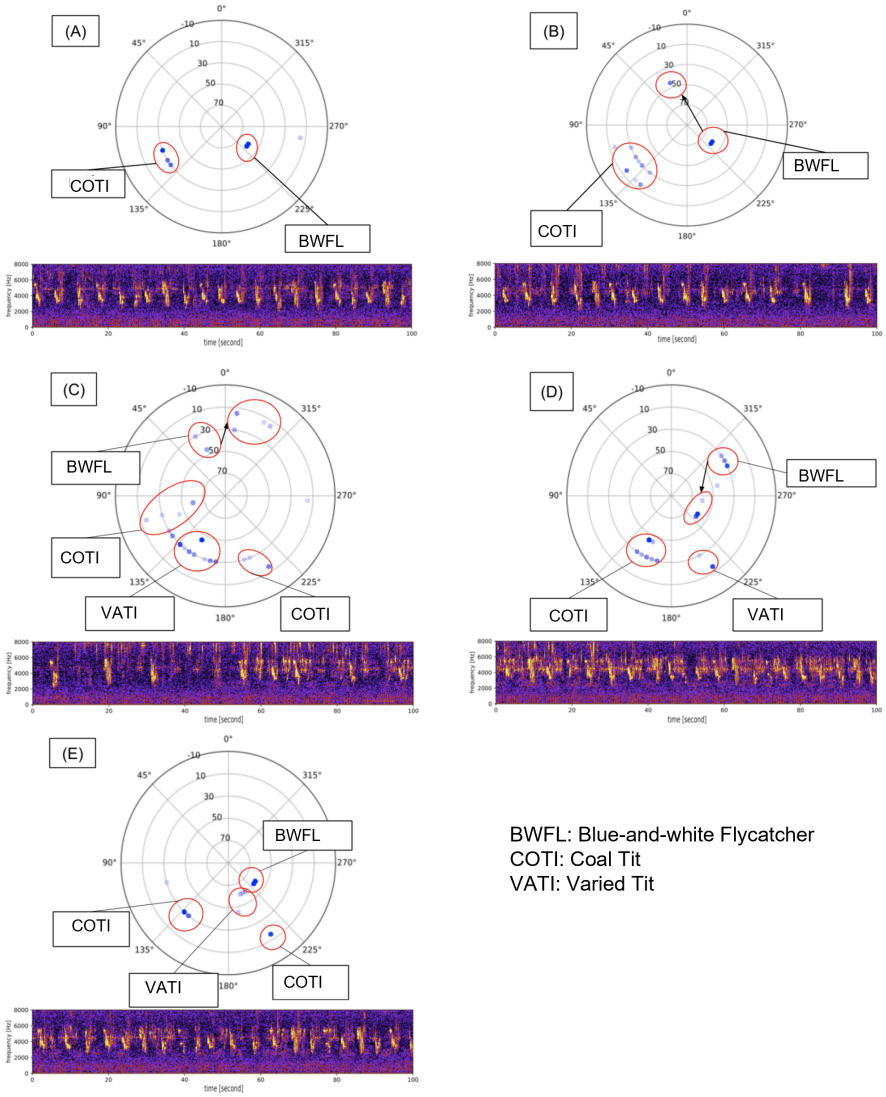

Figure 7. A song distribution on the polar coordinate of azimuth (angle) and elevation (radius) in each time slot (A–E) in Figure 5.

Table 1. The variations in azimuth and elevation of localized sounds in each time slot.

Time Slot (s)	# of Localized Songs	Azimuth Variation (rad^2)	Elevation Variation (rad^2)
A (0–100)	37	0.87	0.07
B (550–650)	31	1.15	0.10
C (650–750)	46	0.81	0.04
D (750–850)	57	0.76	0.06
E (900–1000)	45	0.48	0.09

4. Discussion and Conclusions

We discussed the applicability of robot audition techniques to understand the soundscape dynamics of bird vocalizations in forests. We focused on the elevation information of localized bird songs in addition to azimuth information. A speaker test for DOA esti-

mation of replayed vocalizations showed that the observed DOA of a distant target sound matched well with expected values in both azimuth and elevation angle when no other bird vocalizations were present near the target. A field observation of several individuals reflected well the ecologically plausible structures of the soundscape of bird species in the experimental forest, showing vertical species structures of bird vocalizations. Several statistical indices of localized songs can also summarize the detailed changes in the structures of the soundscape.

The localization of bird vocalization is based on various components including both hardware (i.e., microphone arrays) and software (i.e., HARK). Finer resolution of the estimated DOA would be an important factor for this purpose because many sound sources other than those of the target species or individuals always coexist in fields. Improving the resolution of the MUSIC spectrum by increasing the number of steering vectors (i.e., candidate directions for DOA estimation) would be useful, but requires a great deal of computational cost, especially for azimuth-elevation estimation. The interpolation of the MUSIC spectrum used for finer 2D localization of bird vocalizations with two microphone arrays [22] would be efficient in this case. The balanced settings of DOA resolution along with interpolation would be beneficial for long-term analyses for biodiversity surveys. Further consideration of the effects of microphone channel geometry on the localization accuracy of bird vocalizations is part of our future work.

A systematic comparison of other sound source localization and separation techniques, including adaptive filtering, is important for more practical applications of robot audition techniques to bird behavioral observations. In this study, we employed the simplest and most standard methods (SEVD-MUSIC and GHDSS) employed in HARK, expecting that it will provide a baseline result because the method has been shown to be applicable to field observations of birds in previous studies as introduced in Introduction. We also expected that using such a simple method would be appropriate to examine the basic effects of acoustic noise in the natural environment, and advanced methods can improve the results (e.g., the MUSIC based on generalized singular value decomposition (GSVD-MUSIC) for better speech recognition [23], and the MUSIC method based on incremental generalized eigenvalue decomposition (iGEVD-MUSIC) for drone audition [24]).

Also, this research has an experimental rather than a theoretical aspect. Still, we believe it is important for considering the trends and challenges in robotic applications to show an example of the application of robotics to field observations of natural sounds. At the same time, we believe that a report on sound source localization in both elevation and azimuth angles is particularly important for birds that can fly. The report will contribute to the practical application of related techniques to ecoacoustics, as microphone arrays are expected to be used more frequently in this field.

The spatial localization of bird songs using multiple microphone arrays (i.e., an array of arrays) is a promising approach to determine the precise location of vocalizations. A system with three microphone array units estimated the location of two color banded Great-Reed Warbler's song posts in a reed marsh with a mean error distance of 5.5 m from the location of the observed song posts [18]. Also, various types of animal vocalization systems based on many microphone units deployed over fields have been proposed recently [25]. Gayk et al. successfully 3D triangulated, using a time difference of arrival (TDOA) approach, calls of warblers using a large microphone array unit system in which channels were far apart from each other [26]. However, it could be costly to deploy and calibrate multiple units in field observations. Our approach based on a single but multi-channel array unit, showing good accuracy of azimuth-elevation angles of bird vocalizations, suggest another possibility to better capture bird vocalizations while keeping deployment costs low.

Our results show how our observation method could be used to noninvasively monitor rare birds in the field. For example, Matsubayashi et al. evaluated the practical effectiveness of localization technology for auditory monitoring of endangered Eurasian bittern (*Botaurus stellaris*) which inhabits wetlands in remote areas with thick vegetation, using a 8-ch microphone array unit [27]. They successfully localized booming calls of at least two males

in a reverberant wetland, surrounded by thick vegetation and riparian trees. In addition to the non-invasiveness to the ecosystem where the target birds inhabit, our recording system has lower deployment cost for field observers. We believe that our monitoring system, given advantages and limitations presented in this study, offers a practical tool for field ecologists, e.g., to estimate abundance and distribution of rare species.

However, estimating the distance of sound sources from microphones and two-dimensional (spatial) localization of them are important or essential for more detailed ecological surveys. We think that extracting any complementary information about their distance from separated sounds (e.g., relative amplitude [28]) would be a novel direction to better capture the structure of the soundscape with a single microphone array unit.

From another perspective, there is increasing interest and development of sound event localization and detection for various environmental sounds using microphone arrays. For example, the workshop on detection and classification of acoustic scenes and events (DCASE) provided a dataset (STARSS22) for sound source localization and classification of domestic sounds in indoor environments [29]. A competition of sound source localization and classification has been conducted, and participants discuss issues arising from the task (e.g., [11]). Experimental reports on the sound source localization of distant and elevated calls in a forest environment where many species of birds coexist, which was investigated in this study, could contribute to further progress in these fields because it may provide different insights into sound localization in harsher conditions unique to natural acoustic environments.

Although camera trap-based animal monitoring combined with object detection algorithms is widely used [30], it is challenging to capture small animals, such as songbirds, because they are basically far distant from the device, and there exists the problem of backlighting. This experiment shows that it is possible to quantitatively extract the dynamics of the use of niches among species, which could only be described verbally or roughly before, even when the method is based only on azimuth and elevation angle information.

The increasing interest and popularity of 3D audio in public has made portable 3D recording equipment more accessible (e.g., Zoom H3-VR; Zoom Inc., GoPro MAX; GoPro Inc.). It is worth mentioning that these microphone units (or cameras with multiple microphones) are inexpensive, portable and easily affordable, even with the significant disadvantage of poor sound source localization performance due to their small size. This study also suggests the possibility of using this type of portable and easily available microphone array in ecoacoustics, which can contribute to citizen science of ornithology [31] in addition to the recent development of bird song extraction apps based on deep learning techniques (e.g., BirdNet [32]). One of the problems in the application of these approaches is the low accuracy in detection of vocalizations overlapped with each other or with other environmental sound sources. The robot audition techniques can resolve this problem by separating sound sources by making use of spectogram information from multiple channels, as discussed in this paper.

The future work includes practical comparisons of the efficiency of bird song localization and separation between the microphone arrays adopted in this study and such commercially available ones in order to further explore the applicability of robot audition techniques to ecoacoustic research.

Author Contributions: Conceptualization, K.H., H.O. and R.S.; field experiment K.H. and H.O.; formal analysis and writing, K.H. and R.S., supervision, S.M., T.A., K.N. and H.G.O., funding: R.S. and K.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS/MEXT KAKENHI: JP21K12058, JP20H00475, JP19KK0260.

Institutional Review Board Statement: The experimental procedures have been approved by the planning and evaluation committee in the Graduate School of Information Science, Nagoya University (GSI-H30-1).

Data Availability Statement: The experimental data are available upon request.

Acknowledgments: We thank Noriyoshi Kaneko and Shiro Murahama for designing and conducting a speaker test. We also thank Naoki Takabe (Nagoya Univ.) for conducting a field recording in the experimental forest.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gasc, A.; Francomano, D.; Dunning, J.B.; Pijanowski, B.C. Future directions for soundscape ecology: The importance of ornithological contributions. *Auk* **2016**, *134*, 215–228. [CrossRef]
2. Stowell, D. Computational Bioacoustic Scene Analysis. In *Computational Analysis of Sound Scenes and Events*; Virtanen, T., Plumbley, M.D., Ellis, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; Chapter 11, pp. 303–333.
3. Farina, A.; Gage, S.H. *Ecoacoustics: The Ecological Role of Sounds*; John Wiley and Sons: Hoboken, NJ, USA, 2017.
4. Catchpole, C.K.; Slater, P.J.B. *Bird Song: Biological Themes and Variations*; Cambridge University Press: Cambridge, UK, 2008.
5. Goëau, H.; Gloti, H.; Vellinga, W.P.; Planqué, R.; Joly, A. LifeCLEF Bird Identification Task 2016: The arrival of Deep learning. In Proceedings of the CLEF: Conference and Labs of the Evaluation Forum, Évora, Portugal, 5–8 September 2016.
6. Suzuki, R.; Cody, M.L. Complex systems approaches to temporal soundspace partitioning in bird communities as a self-organizing phenomenon based on behavioral plasticity. *Artif. Life Robot.* **2019**, *24*, 439–444. [CrossRef]
7. Rhinehart, T.A.; Chronister, L.M.; Devlin, T.; Kitzes, J. Acoustic localization of terrestrial wildlife: Current practices and future opportunities. *Ecol. Evol.* **2020**, *10*, 6794–6818. [CrossRef] [PubMed]
8. Harlow, Z.; Collier, T.; Burkholder, V.; Taylor, C.E. Acoustic 3D localization of a tropical songbird. In Proceedings of the IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP), Beijing, China, 6–10 July 2013; pp. 220–224.
9. Hedley, R.W.; Huang, Y.; Yao, K. Direction-of-arrival estimation of animal vocalizations for monitoring animal behavior and improving estimates of abundance. *Avian Conserv. Ecol.* **2017**, *12*, 6. [CrossRef]
10. Gabriel, D.; Kojima, R.; Hoshiba, K.; Itoyama, K.; Nishida, K.; Nakadai, K. Case study of bird localization via sound in 3D space. In Proceedings of the 36th Annual Conference of the Robotics Society of Japan, Tokyo, Japan, 21 August 2018; p. RSJ2018AC1I2–06.
11. Nguyen, T.N.T.; Watcharasupat, K.N.; Lee, Z.J.; Nguyen, N.K.; Jones, D.L.; Gan, W.S. What makes sound event localization and detection difficult? Insights from error analysis. In Proceedings of the 6th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE 2021), online, 15–19 November 2021; pp. 120–124.
12. Nakadai, K.; Okuno, H.G.; Mizumoto, T. Development, Deployment and Applications of Robot Audition Open Source Software HARK. *J. Robot. Mechatronics* **2017**, *27*, 16–25. [CrossRef]
13. Sumitani, S.; Suzuki, R.; Matsubayashi, S.; Arita, T.; Nakadai, K.; Okuno, H.G. An integrated framework for field recording, localization, classification and annotation of birdsongs using robot audition techniques - HARKBird 2.0. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Brighton, UK, 12–17 May 2019; pp. 8246–8250.
14. Suzuki, R.; Sumitani, S.; Naren, N.; Matsubayashi, S.; Arita, T.; Nakadai, K.; Okuno, H.G. Field observations of ecoacoustic dynamics of a Japanese bush warbler using an open-source software for robot audition HARK. *J. Ecoacoustics* **2018**, *2*, EYAJ46. [CrossRef]
15. Sumitani, S.; Suzuki, R.; Matsubayashi, S.; Arita, T.; Nakadai, K.; Okuno, H.G. Fine-scale observations of spatio-spectro-temporal dynamics of bird vocalizations using robot audition techniques. *Remote Sens. Ecol. Conserv.* **2020**, *7*, 18–35. [CrossRef]
16. Huang, Q.; Swatantran, A.; Dubayah, R.; Goetz, S.J. The Influence of Vegetation Height Heterogeneity on Forest and Woodland Bird Species Richness across the United States. *PLoS ONE* **2014**, *9*, 103236. [CrossRef]
17. Matsubayashi, S.; Saito, F.; Suzuki, R.; Matsubayashi, S.; Arita, T.; Nakadai, K.; Okuno, H.G. Observing Nocturnal Birds Using Localization Techniques. In Proceedings of the 2021 IEEE/SICE International Symposium on System Integrations (SII), Virtual, 11–14 January 2021; pp. 493–498.
18. Suzuki, R.; Matsubayashi, S.; Saito, F.; Murate, T.; Masuda, T.; Yamamoto, Y.; Kojima, R.; Nakadai, K.; Okuno, H.G. A Spatiotemporal Analysis of Acoustic Interactions between Great Reed Warblers (Acrocephalus arundinaceus) Using Microphone Arrays and Robot Audition Software HARK. *Ecol. Evol.* **2018**, *8*, 812–825. [CrossRef]
19. Schmidt, R. Bayesian Nonparametrics for Microphone Array Processing. *IEEE Trans. Antennas Propag. (TAP)* **1986**, *34*, 276–280. [CrossRef]
20. Nakajima, H.; Nakadai, K.; Hasegawa, Y.; Tsujino, H. Blind source separation with parameter-free adaptive step-size method for robot audition. *IEEE Trans. Audio Speech Lang. Process.* **2010**, *18*, 1476–1484. [CrossRef]
21. Suzuki, R.; Matsubayashi, S.; Nakadai, K.; Okuno, H.G. HARKBird: Exploring acoustic interactions in bird communities using a microphone array. *J. Robot. Mechatronics* **2017**, *27*, 213–223. [CrossRef]
22. Sumitani, S.; Suzuki, R.; Matsubayashi, S.; Arita, T.; Nakadai, K.; Okuno, H.G. Extracting the relationship between the spatial distribution and types of bird vocalizations using robot audition system HARK. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, Spain, 1–5 October 2018; pp. 2485–2490.
23. Nakamura, K.; Nakadai, K.; Okuno, H.G. A real-time super-resolution robot audition system that improves the robustness of simultaneous speech recognition. *Adv. Robot.* **2013**, *27*, 933–945. [CrossRef]

24. Okutani, K.; Yoshida, T.; Nakamura, K.; Nakadai, K. Outdoor Auditory Scene Analysis Using a Moving Microphone Array Embedded in a Quadrocopter. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012), Algarve, Portugal, 7–12 October 2012; pp. 3288–3293.
25. Verreycken, E.; Simon, R.; Quirk-Royal, B.; Daems, W.; Barber, J.; Steckel, J. Bio-acoustic tracking and localization using heterogeneous, scalable microphone arrays. *Commun. Biol.* **2021**, *4*, 1275. [CrossRef] [PubMed]
26. Gayk, Z.; Mennill, D.J. Pinpointing the position of flying songbirds with a wireless microphone array: Three-dimensional triangulation of warblers on the wing. *Bioacoustics* **2019**, *29*, 375–386. [CrossRef]
27. Matsubayashi, S.; Nakadai, K.; Suzuki, R.; Ura, T.; Hasebe, M.; Okuno, H.G. Auditory Survey of Endangered Eurasian Bittern Using Microphone Arrays and Robot Audition. *Front. Robot. AI* **2022**, *9*, 854572. [CrossRef] [PubMed]
28. Hedley, R.W.; Wilson, S.J.; Yip, D.A.; Li, K.; Bayne, E.M. Distance truncation via sound level for bioacoustic surveys in patchy habitat. *Bioacoustics* **2021**, *30*, 303–323. [CrossRef]
29. Politis, A.; Shimada, K.; Sudarsanam, P.; Adavanne, S.; Krause, D.; Koyama, Y.; Takahashi, N.; Takahashi, S.; Mitsufuji, Y.; Virtanen, T. STARSS22: A dataset of spatial recordings of real scenes with spatiotemporal annotations of sound events. *arXiv* **2022**, arXiv:2206.01948.
30. Tan, M.; Chao, W.; Cheng, J.K.; Zhou, M.; Ma, Y.; Jiang, X.; Ge, J.; Yu, L.; Feng, L. Animal Detection and Classification from Camera Trap Images Using Different Mainstream Object Detection Architectures. *Anim. Open Access J. MDPI* **2022**, *12*, 1976. [CrossRef] [PubMed]
31. Tulloch, A.I.; Possingham, H.P.; Joseph, L.N.; Szabo, J.; Martin, T.G. Realising the full potential of citizen science monitoring programs. *Biol. Conserv.* **2013**, *165*, 128–138. [CrossRef]
32. Wood, C.M.; Kahl, S.; Rahaman, A.; Klinck, H. The machine learning–powered BirdNET App reduces barriers to global bird research by enabling citizen science participation. *PLoS Biol.* **2022**, *20*, e3001670. [CrossRef] [PubMed]