Paper:

HARKBird: Exploring Acoustic Interactions in Bird Communities Using a Microphone Array

Reiji Suzuki[*1], **Shiho Matsubayashi**[*1], **Richard W. Hedley**[*2], **Kazuhiro Nakadai**[*3,*4], **and Hiroshi G. Okuno**[*5]

[*1]Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
E-mail: reiji@nagoya-u.jp
[*2]Department of Ecology and Evolutionary Biology, University of California Los Angeles
Los Angeles, CA 90095, USA
[*3]Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
[*4]Honda Research Institute Japan Co., Ltd.
8-1 Honcho, Wako, Saitama 351-0114, Japan
[*5]Graduate School of Fundamental Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
[Received August 1, 2016; accepted October 12, 2016]

Understanding auditory scenes is important when deploying intelligent robots and systems in real-world environments. We believe that robot audition can better recognize acoustic events in the field as compared to conventional methods such as human observation or recording using single-channel microphone array. We are particularly interested in acoustic interactions among songbirds. Birds do not always vocalize at random, for example, but may instead divide a soundscape so that they avoid overlapping their songs with those of other birds. To understand such complex interaction processes, we must collect much spatiotemporal data in which multiple individuals and species are singing simultaneously. However, it is costly and difficult to annotate many or long recorded tracks manually to detect their interactions. In order to solve this problem, we are developing HARKBird, an easily-available and portable system consisting of a laptop PC with open-source software for robot audition HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) together with a low-cost and commercially available microphone array. HARKBird enables us to extract the songs of multiple individuals from recordings automatically. In this paper, we introduce the current status of our project and report preliminary results of recording experiments in two different types of forests – one in the USA and the other in Japan – using this system to automatically estimate the direction of arrival of the songs of multiple birds, and separate them from the recordings. We also discuss asymmetries among species in terms of their tendency to partition temporal resources.

Keywords: bird songs, localization, temporal soundscape partitioning, microphone array, HARK

1. Introduction

Understanding auditory scenes is important when deploying intelligent robots and systems in real-world environments. Sound information, however, has not been so widely utilized compared to visual information in environmental monitoring and management. We believe that recent advances in signal processing for robot audition enable robotic systems to recognize various acoustic events in natural habitats better than conventional single-channel recording or human observation.

In ornithology or bird observation, songs of birds provide critical cues for monitoring their behavior. In forests, many male birds produce long vocalizations, called songs, to advertise their territory or attract females in breeding season. Songs vary by species that could consist of multiple phrases. Conversely, shorter vocalizations occurs in other contexts such as flight, threat, and alarm [1]. In general, a species can be identified by songs more readily than by calls, but this strongly depends on species-specific properties.

A community of singing birds can be regarded as a self-organizing system in the sense that they establish an efficient soundscape through acoustic interactions with neighboring individuals. Rather than vocalizing at random, for example, birds may divide a soundscape so that they avoid overlapping their songs with songs of other bird species or individuals to communicate with neighbors efficiently. Empirical studies on temporal partitioning or overlap avoidance of singing behaviors of songbirds have been conducted across various time scales [2–8]. We are interested in clarifying underlying dynamics

as an example of such complex systems based on adaptive behavioral plasticity from both theoretical [9] and empirical standpoints [10, 11], which might share general properties with other biological (e.g., insects, reptiles [12]) and engineering (e.g., sensor networks [13]) systems.

Much spatiotemporal data in which multiple individuals and species are singing simultaneously must be collected to understand such complex interaction processes. However, there are various issues for this task based on a standard single-channel recording: 1) we have to perform long-term recordings because we do not know when target species or individuals sing simultaneously. 2) It is costly or even impossible to annotate manually such long tracks because songs could be intermixed in single-channel recordings. 3) No spatial or directional information is available, which will be a key for distinguishing individuals.

Using a microphone array for acoustic monitoring of animals is a promising approach [14]. Collier et al. developed a wireless network of 8 nodes of 4-ch microphone arrays (VoxNet) and showed that songs of wild birds such as Mexican Antthrush (*Formicarius moniliger*) can be spatially localized successfully [15]. Harlow et al. expanded this system to 3D-localization of Shrike-tanagers (*Lanio aurantius*) and Wood Wrens (*Henicorhina leucosticta*) in a Mexican rainforest [16].

Despite the great potential of these techniques to enable us to better understand how birds communicate via songs, systems developed in such studies are often not commonly available or may only available upon request. This prevents many field researchers from making use of the latest technologies.

In addition, robustness against noises is an important factor in field observations. Mennill et al. constructed an array of multiple commercial stereo recorders (Songmeter SM2 with GPS; Wildlife Acoustics Inc.) [17]. Recorded sounds were synchronized to generate 8-channel data and bird or animal calls were extracted manually. The 2D location of each source was estimated based on a cross-correlation method [18] in MATLAB. This system showed a high accuracy in localizing a variety of sounds, including bird songs replayed by a loud speaker, under ideal conditions in which a single target sound source was played in a relatively quiet environment. In contrast, we aim to grasp a more realistic representation of the soundscape in which multiple individuals or species sing simultaneously in noisy environments.

To this end, we are developing an easily-available and portable system called HARKBird [19]. HARKBird consists of a standard laptop PC with an open source software for robot audition HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) [20] and a low-cost and commercially available microphone array. This enables us to automatically extract songs of multiple individuals or species from recordings.

While some systems consisting of commercially available microphones for research use have been proposed [17], they are still expensive and require expertise for practical implementation. Some software packages [21] have also been proposed with a similar motivation, but one significant benefit of HARK is its continuous updating since its original release in 2010 to include the latest algorithms for sound source localization, separation, and recognition. A MUSIC (Multiple Signal Classification) method, which is adopted as a sound source localization algorithm in HARK, realizes noise-robust and high-resolution localization of multiple sound sources under conditions where the number of sources is fewer than the number of microphone elements. These features are well suited for analyzing complex acoustic environments in natural habitats.

We are using a single microphone array, which means that we can only estimate the arrival direction of sound sources rather than the spatial location. Even in such a minimal case, rich acoustic information not available from a single-channel recording enables us to grasp the overall soundscape of bird songs in detail. That is, a microphone array reduces annotation cost while increasing accuracy, especially where multiple individuals or species are singing.

HARKBird also has the benefits of being portable and customizable. We can see localization results immediately after recording in the field. The customizability of HARK also enables us to make the system respond to localized acoustic events in real-time.

Our purpose in this work is twofold: first, to introduce our system; and second, to show some localization results of bird songs recorded in the USA and Japan, in order to discuss advantages and limitations of HARKBird. We also discuss asymmetric interactions among species that engaged in temporal resource partitioning in the recording in Japan.

2. HARKBird: A Portable System for Recording, Localizing and Annotating Bird Songs

HARKBird is a portable system we developed for recording, localizing[1] and annotating the singing behaviors of multiple songbirds. HARKBird consists of a laptop PC and a commercially available USB microphone array. **Fig. 1** shows an overview of the system. **Fig. 1(a)** shows the system in the field. In our pilot trials here, we used a TOUGHBOOK CF-C2/CF-19 (Panasonic) and a Microcone (Dev-Audio),[2] a 7-channel microphone array on a tripod. The Microcone has a small cone-shaped body which is 10 cm high and 8 cm in bottom diameter (**Fig. 1(b)**). It has 6 microphones 60° apart on the bottom and 1 microphone on top.

For the software environment, we adopted Ubuntu Linux 12.04 in which the latest HARK and HARK-Python were installed. The whole system consists of HARK and a set of Python scripts with major modules (e.g., wxpython and pyside) and standard software for

1. In this paper, we use the term "localize" as an estimate of the direction of arrival in 2D without distance information.
2. Microcone is discontinued. Instead, TAMAGO, a low-price 8-channel USB microphone array, is available from System In Frontier (http://www.sifi.co.jp/en/ [Accessed January 30, 2017]).

Fig. 1. An overview of the system. (a) The system in the field. (b) A Microcone, which has a small cone-shaped body which is 10 cm high and 8 cm in bottom diameter. It has 6 microphones 60° apart on the bottom and 1 microphone on top. (c) GUI interface. (d) A network of HARK for localizing sound sources in a wave file with MUSIC (Multiple Signal Classification) using 7ch spectrograms with FFT, then separating localized sounds with GHDSS (Geometric High order Decorrelation-based Source Separation) method.

sound processing (e.g., sox, arecord and aplay). The scripts are available on our website.[3]

We introduce functions of the system that enable us to record, localize, visualize and annotate bird songs. All components of HARKBird are primitive but essential to record and annotate bird songs in the field such as a forest.

2.1. Recording

Figure 1(c) shows the GUI interface of the system. We can start and stop 7ch-recording (16 bit and 16/48 kHz) quickly with this interface. Although online processing is a significant benefit of HARK, we decided to record offline to maximize recording duration by minimizing the computational cost of the system on a battery-powered laptop PC. In our trial recordings with a Panasonic TOUGHBOOK CF-C2 with a large battery, it lasted over 8 hours, which is enough to record bird songs from dawn to noon. The name of each recording can be assigned according to the staring time, the date and a unique ID. This enables us to start recording without overwriting recorded files, which is a simple but essential requirement for field research to avoid losing changes for recording acoustic interactions among target species.

2.2. Localization and Separation Using HARK

We can start sound source localization and separation of recorded files using HARK by choosing a recorded wave file from a list of wave files created in the current directory. **Fig. 1(d)** shows the network we used for sound localization and separation. This network reads a recorded wave file down converted to 16 kHz, localizes sound sources with the MUSIC (Multiple Signal Classification) method using 7-ch spectrograms with the FFT, and separates localized sounds with the GHDSS (Geometric High order Decorrelation-based Source Separation) method. We can specify some parameters for the MUSIC method and source tracking that are important for localizing bird songs successfully as follows:

1. The expected number of sound sources for the MUSIC method: Three or four sound sources are appropriate for tracking singing behavior of some species in our study sites, although this depends on the time and place.
2. The lower bound frequency for the MUSIC method:

3. http://www.alife.cs.is.nagoya-u.ac.jp/~reiji/HARKBird/ [Accessed January 30, 2017]

In forests, most noise originates from leaves, water, and wind. On the other hand, the frequency range of songs of major species is higher than such low-frequency noise. Thus we usually set value at 1.5–2.5 kHz for localization processes, although this value depends on the situation because we may fail to localize the songs of species that have lower-frequency songs. We fixed the upper bound frequency at 10 kHz, which covers the frequency range of songs of major species.

3. The threshold for source tracking: This parameter determines a power threshold, and the system assesses by the threshold whether the source localization result is a noise to be ignored. Apparently, the appropriate value for this parameter strongly depends on the acoustic environment of a recording, and thus it is important to be able to adjust it readily at the time of analysis.

Based on localization and separation, the following files will be created in a new folder: separated wave files, log files of localized sources and a MUSIC spectrum at each time step, lists of information on localized sounds (start time, the direction of arrival (DOA) and duration) in JSON format.

2.3. Visualization

Using the exported files above, HARKBird provides several ways to visualize and analyze the acoustic environment of recording. The GUI interface displays interactive windows containing spectrograms and localization results, in which the duration and the DOA of each sound is represented as a line. Each separated sound can be replayed and visualized with a spectrogram by clicking on the line ID, as shown in **Fig. 2**. HARKBird generates a PDF file of localization results that include the spectrogram of the recording, the MUSIC spectrum, and the directional and temporal patterns of sound localization as shown in **Fig. 3**. This PDF file is useful for visualizing the long-term patterns of the acoustic environment.

2.4. Annotation

Our research group on ecological analyses of bird songs[4] uses an open-source software for human speech analysis "Praat"[5] for annotating the timing, duration and types of songs or phrases in tracks recorded in California. A tier, a timeline for representing the timing and duration of songs, is assigned to each species or individual as shown in **Fig. 4(b)**. Before annotating results with Praat, noise should be excluded and localization results of the same individual should be grouped into individual birds. In this grouping, we assume that they are singing periodically without moving around drastically. Results can be exported in standard Praat format or as a figure (**Fig. 5**).

4. http://artsci.ucla.edu/birds/ [Accessed January 30, 2017]
5. http://www.fon.hum.uva.nl/praat/ [Accessed January 30, 2017]

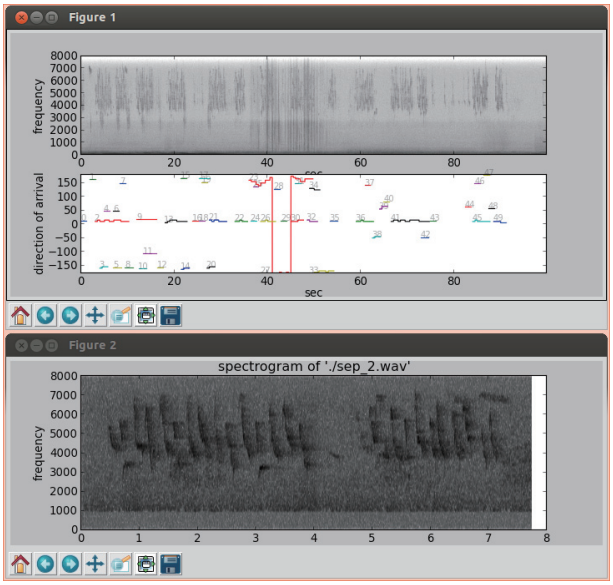

Fig. 2. Example of interactive windows of localization results. The track was recorded at Higashiyama-park, Nagoya, Japan (April 2014). A Japanese White-eye (*Zosterops japonicus*) singing repeatedly at 10° and a Brown-eared Bulbul (*Hypsipetes amaurotis*) singing repeatedly at −150°.

As another approach, we provide a simple and minimal annotation tool for editing and classifying localization results, as shown in **Fig. 6(b)**, which will be discussed in detail (Section 3.2.1). We can load a wave file and corresponding files of localization results. The tool overlays the temporal and directional distribution of localized sound sources in the space of MUSIC spectrum. This tool has a minimal interface for correcting the timing, direction and duration of individual localized sounds, to add a new source that has not been localized by HARK, to remove unnecessary sources, and to assign labels to sources. The modified results can be saved in JSON format.

3. Two Case Studies

We introduce two cases of localization analysis, and discuss the advantages and limitations of the current HARKBird. We expect that the minimal set of analysis tools and the GUI of HARKBird will greatly assist users in collecting acoustic data in the field. These features will be particularly useful for bird researchers who are not necessarily familiar with operating systems and localization techniques.

3.1. Example of Recording in the USA

3.1.1. Localization of Different Types of Bird Vocalizations

We conducted several pilot recordings at our field site in a mixed conifer-oak forest near Volcano, CA, USA (May 2015). **Fig. 3** shows localization results for a recording of approximately four minutes. The local-

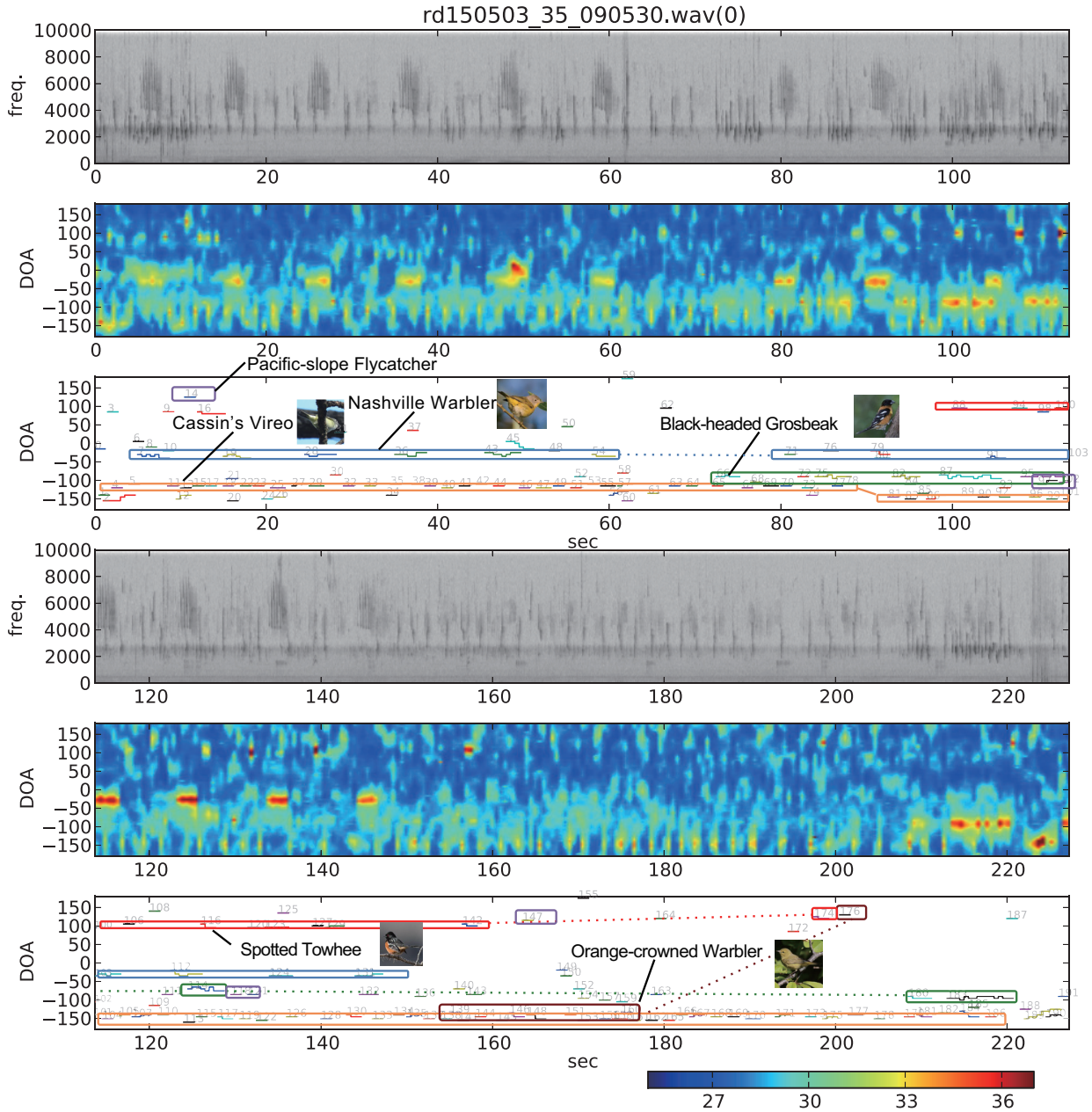

Fig. 3. An example localization of a recording at our field site in a mixed conifer-oak forest near Volcano, CA, USA (May 2015). We used the following parameter settings in Section 2.1: 1) 3 sources, 2) 2400 Hz and 3) 29.5. The classification of species was conducted manually. Each rectangle roughly indicates the spatiotemporal pattern of the vocalizations of the corresponding species, thus it may contain noise and the songs or calls of the focal species. NAWA: Nashville Warbler (*Leiothlypis ruficapilla*), BHGR: Black-headed Grosbeak (*Pheucticus melanocephalus*), OCWA: Orange-crowned Warbler (*Vermivora celata*), CAVI: Cassin's Vireo (*Vireo cassinii*), PSFL: Pacific-slope Flycatcher (*Empidonax difficilis*), SPTO: Spotted Towhee (*Pipilo maculatus*).

ization results were based on all seven microphone elements. We focus on songs and calls of six species because they were recognized at least five times each: the Nashville Warbler (*Leiothlypis ruficapilla*, NAWA); the Black-headed Grosbeak (*Pheucticus melanocephalus*, BHGR:); the Orange-crowned Warbler (*Vermivora celata*, OCWA); the Cassin's Vireo (*Vireo cassinii*, CAVI); the Pacific-slope Flycatcher (*Empidonax difficilis*, PSFL); and the Spotted Towhee (*Pipilo maculatus*, SPTO).

First, we see that sound sources were repeatedly localized at fixed degrees of arrival. This fact means that several individuals were singing at these directions periodically. By replaying the separated sound or comparing the timing of localized sounds with the spectrogram for the corresponding period, we aurally confirmed which species was singing at each direction, as shown in **Fig. 3**.

For example, several sound sources localized periodically at $-30°$ are the songs of the NAWA. A song of the CAVI consists of different types of short phrases with a short break between phrases localized at around $-120°$ and $-150°$ throughout the recording. This fluctuation shows that it slightly moved from $-120°$ to $-150°$. Thus,

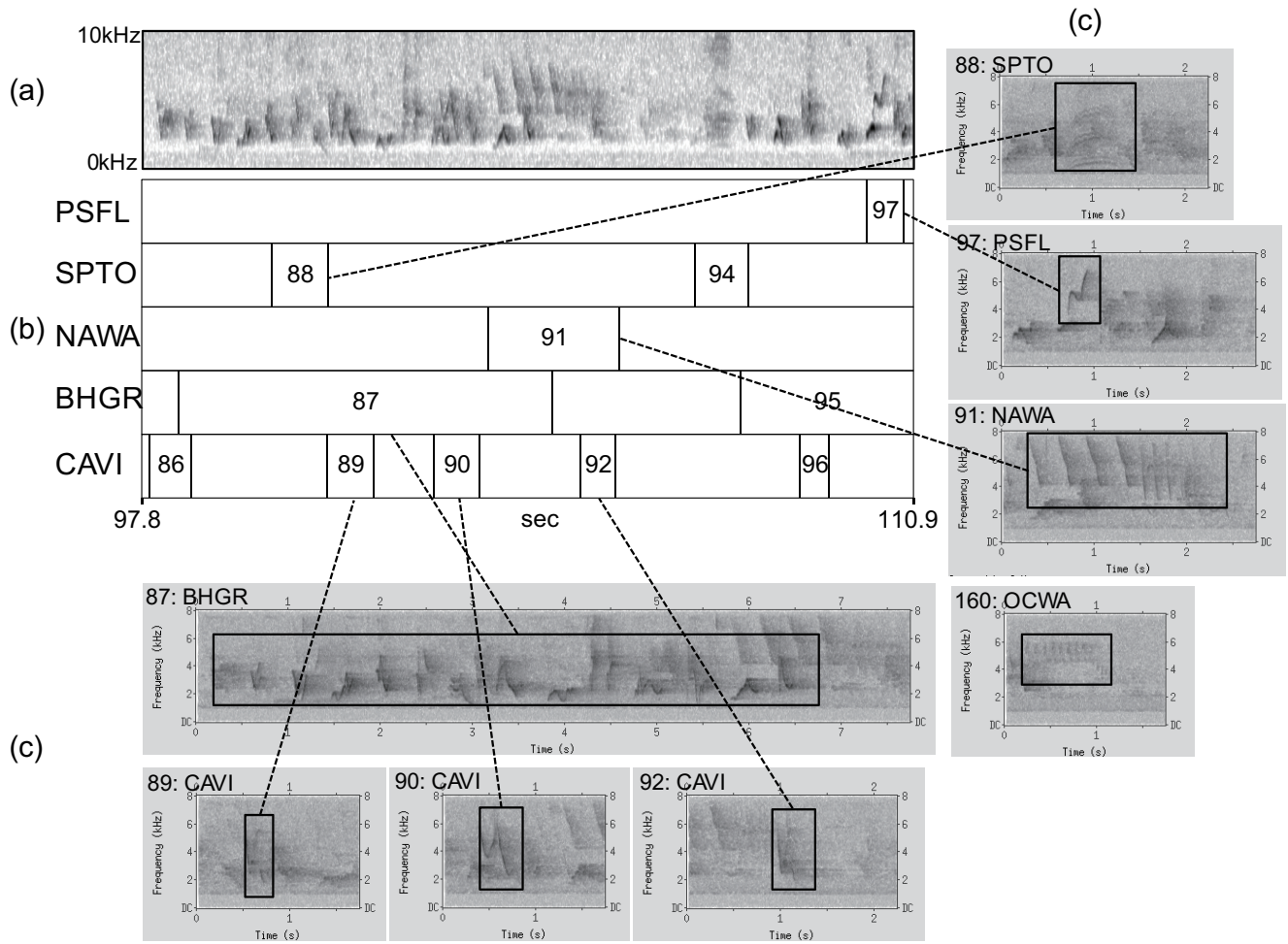

Fig. 4. Detailed results of localization at the duration between 97.8 and 110.9 seconds in **Fig. 3**. (a) The spectrogram of the original recording. (b) Tiers (timelines) of songs for each species. (c) Spectrograms of separated songs.

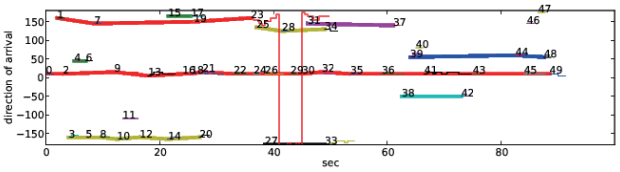

Fig. 5. Example of ad-hoc classification of localized sources in **Fig. 2**. The sources categorized into the same class are connected with thick lines.

localization results indicate that this soundscape consisted of songs of different bird species.

These species have the distinct, species-specific properties in their vocalizations. We focused on the interval from 97.8 to 110.9 seconds in **Fig. 4**. The spectrogram of the first channel of the original recording in **Fig. 4(a)** shows songs of the different species. The tiers (timelines) of Praat in **Fig. 4(b)** show the timing and duration of separated sources for each species, together with their IDs. In the spectrograms of some separated vocalizations of these species, such as in **Fig. 4(c)**, a rectangular region on the spectrogram represents the duration and frequency range of focal songs, phrases or calls. These vocalizations are observed more clearly in **Fig. 4(c)** than **Fig. 4(a)** thanks to the successful separation of sound sources. However, the durations of the separated sources tend to be longer than the target vocalizations, possibly caused by other songs, especially those of the BHGR, which has a longer and louder song. The source with the ID 91 corresponds to the song of the NAWA, which is slightly longer (2.5 seconds) and with a higher frequency (at 3–8 kHz). The short sources (ID 86, 89, 90, 92 and 96) are phrases of the CAVI, whose song consists of different types of short phrases with a short break between phrases. On the other hand, many phrases of the BHGR were localized as a single and long source (87) because intervals between phrases were too short to be localized as separate sources. The parameters of HARK can, of course, control the minimum pause length.

Without HARKBird, we could not recognize the call of the SPTO (88) by neither looking at the spectrograms nor listening to the recording because it was singing too faintly, possibly singing faraway. Nevertheless, we recognize the species-specific property of a call of the SPTO in the spectrogram of the localized source (88), and recog-

Table 1. Accuracy of localization.

Species type	NAWA song	BHGR song	OCWA song	CAVI phrase	PSFL call	SPTO call
Actual song	13	16	10	95	5	12
Localized song	13	12	4	89	4	11
Success rate	100.0	75.0	40.0	93.7	80.0	91.7

nize the call by replaying the separated source. This is one benefit of using HARK for annotation, because it enables us to look closer at sources that are not easily recognized from a single channel recording. Sources 106 and 116 in **Fig. 3** include the similar songs to that of source 88, thus we clearly recognize the existence of this individual at about 100°, vocalizing at regular intervals. Although the call of the PSFL (97) was very short, we recognized it while the BHGR was singing simultaneously. The songs of the OCWA were localized only a few times because this individual appeared to be singing faraway. The localization of this species might have been affected by overlap with other species' songs (e.g., BHGR).

3.1.2. Accuracy Evaluation

To evaluate overall localization accuracy, we conducted fine-grained annotation of this recording by human annotators, finding that some songs or calls were not localized by HARK. Because we do not know the ground truth of the singing behaviors of these species, we defined the success rate of localization for each species as "the ratio of the number of localized songs by HARK to that of actual songs recognized by human annotators or HARK." The ratios are shown in **Table 1**. More than 70% of the songs and calls were localized successfully, except for the OCWA that sang from a distant location. Because the total number of localized sources was 191 and that of actual songs is 133, 30.3% of the localized sources were either noises or a mixture of several species' songs. This result indicates a need for the development of the automatic annotation system, particularly its ability to deal with noise.

3.2. Example of Recording in Japan

3.2.1. Effects of Surrounding Environments on Localization of Bird Songs

Another example is a track recorded at the Inabu field, the experimental forest of the Field Science Center, Graduate School of Bioagricultural Sciences, Nagoya University, in central Japan (May 2015). The forest is mainly 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 the breeding season. **Fig. 6(a)** shows an example of localization result for about 10 min. We focus on the five major species that sang repeatedly in the recording: Blue-and-white Flycatcher (*Cyanoptila cyanomelana*, BAWF), Narcissus Flycatcher (*Ficedula narcissina*, NAFL), Coal Tit (two individuals, *Periparus ater insularis*, COTI), Japanese Bush-Warbler (*Horornis diphone*, JBWA) and Eurasian Wren (*Troglodytes troglodytes*, EUWR).

Several sequences of localized sources were manually classified as indicated in **Fig. 6(a)**. For example, the COTI, NAFL and BAWF were singing at about 120°, 80° and −170° from the microphone at 200 seconds, respectively. The EUWR sang long and loud songs a few times from about −120° at 270 seconds, and the JBWA sang at around −180° during the last 70 seconds. This means that HARKBird can roughly capture the soundscape of singing behaviors of these species. It should be noted that two individuals of the COTI sang at close but slightly different directions. Such directional information of sources was informative to discriminate between different individuals of the same species. Each separated sound was also important for listening to the song more clearly, compared with the original recording. This observation cannot be obtained with a single-channel recording.

This analysis showed several technical limitations of our system. It localized many non-bird song sound sources because the localization results strongly reflected the acoustic environment around the microphone. **Fig. 6(b)** shows a correctly annotated track using the annotation tool to edit results from HARK. Because all noises were removed manually in **Fig. 6(b)**, a comparison between **Figs. 6(a)** and **(b)** shows there were many noises between −150° and 0°. In **Fig. 6(a)**, in particular, there are sound sources localized repeatedly in the direction of 0°, and it turned out that these were not sound sources from real birds but reflected songs of the species mainly at around −170° and −180°, i.e., the BAWF and JBWA reflected by the wall of an old prefabricated hut located at about 0°.

Short sound sources localized between −150° and −50° were continuous noise made by water flow or leaves in a stand of bamboos stand near the microphone. Actually, the values of the MUSIC spectrum in these directions in **Fig. 6(b)** tended to be higher throughout the recording. These effects of the surrounding environments (e.g., obstacles and vegetations) will need to be taken into account in the future. For example, we can remove these unnecessary sounds by using a directional filter, which may improve the overall quality of the localization processes.

3.2.2. Temporal Overlap Avoidance

To see if there was significant temporal overlap avoidance of singing behaviors among the individuals that were

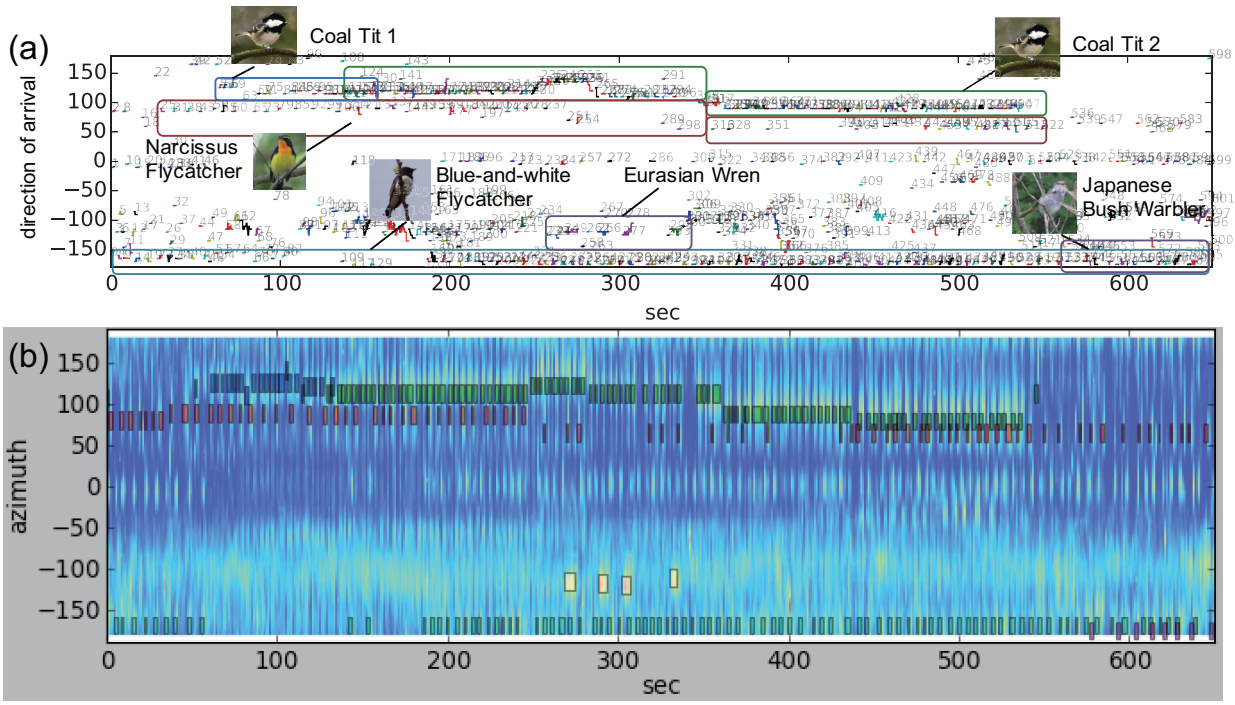

Fig. 6. An example localization of a recording at the Inabu field, the experimental forest of Nagoya University, in central Japan (May 2015). (a) A result of localization by HARK. We used the following the parameter settings in Section 2.1: 1) 3 sources, 2) 2200 Hz and 3) 29.5. (b) A snapshot of the correctly annotated track using the annotation tool by editing the results from HARK. Each rectangle corresponds to the individual in (a). All noises were removed and non-localized songs were added manually. The classification of species was also conducted manually.

actively singing in the recording track between 140 and 583 seconds, we focused on the song timing and duration of COTI (individual 2), NAFL, and BWFL. The duration of time occupied by no species (vacant), monopolized by the songs of a single species, or occupied by the songs of 2 or 3 species were 108, 235 and 100 seconds, respectively. We found that the solo singing time (235 seconds) was significantly longer than would be expected if the birds were singing with random timing (t-test, $p < 0.001$),[6] meaning that temporal overlap avoidance occurred in this time period.

Studies have pointed out that differences exist between species in the tendency to avoid temporal overlaps [2–5]. For example, Suzuki and Arita showed that, using a computational coevolutionary model of a resource sharing problem, inter-specific diversity in behavioral plasticity can emerge and contribute to the efficient and equal benefit of interacting species [9].

To see whether there is such inter-specific diversity in behavioral plasticity during this time, we measured the information flow from one species' behavior to another using transfer entropy (see [22] for detail). This measure quantifies the expected amount of directional information flow from one time series to another.

Specifically, transfer entropy $T_{Y\to X}(k,l)$ from a discrete time series $Y_t = \{y_t\}_{t=1,2,\ldots}$ to another discrete time series $X_t = \{x_t\}_{t=1,2,\ldots}$ represents, given the past l values of Y_t, the amount of reduction in the uncertainty about the future value of Y_t (i.e., the reduced entropy of the transition probability of Y_t) by knowing the past k values of X_t, calculated as follows:

$$T_{Y\to X}(k,l) = \sum \log \frac{p\left(x_{t+1}|x_t^k, y_t^l\right)}{p\left(x_{t+1}|x_t^k\right)}, \quad \ldots\ldots \quad (1)$$

where x_t^k and y_t^l denote $\{x_{t-k+1},\ldots,x_t\}$ and $\{y_{t-l+1},\ldots,y_t\}$, respectively. In our case, X and Y correspond to the time series of singing behavior of the species X and Y when we calculate the information flow from the species X to Y. To discretize each time series, we created a binary time series by assigning a binary value (1: singing or 0: not singing) to each 0.5-second time interval.

Because transfer entropy can be positive even when the states of the additional time series (Y_t) are completely random, the effective transfer entropy $ET_{Y\to X}$ [23] must be calculated:

$$ET_{Y\to X} = T_{Y\to X} - mean(T_{Y_{rand}\to X}), \quad \ldots\ldots \quad (2)$$

where $T_{Y_{rand}\to X}$ is transfer entropy from surrogate time series Y_t, in which durations of two randomly selected non-singing intervals were swapped 1000 times for each species, to observed time series X_t. We generated 50 dif-

6. For the statistical analysis, we created 50 surrogate datasets of singing behaviors in which the durations of two randomly selected non-singing intervals were swapped 1000 times for each species. We then checked whether the observed transfer entropy was significantly different from the average value of those of the surrogate data or not, using a t-test.

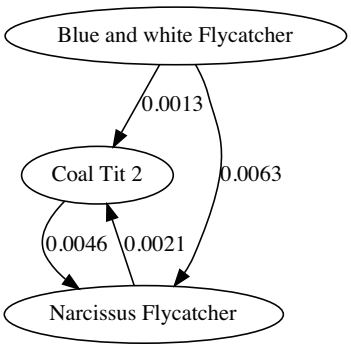

Fig. 7. The network of effective transfer entropy generated from the song timing and durations of COTI 2, NAFL and BWFL between 140 and 582 seconds in **Fig. 6**. Each arrow represents a statistically significant information flow from the species of the outgoing node to that of the incoming node.

ferent $T_{Y_{rand} \to X}$ and used its mean to calculate $ET_{Y \to X}$.[7] Using a t-test ($p < 0.05$), we also tested whether the observed transfer entropy differed significantly from the average of those of 50 surrogate data.

Figure 7 shows a network of statistically significant effective transfer entropy. Each node represents a species and each directional arrow from a node of one species to another represents the existence of statistically significant information flow from the former species to the latter compared to the expected values of surrogate data.

We, therefore, conclude that asymmetric information existed among species. The BAWF, for example, does not receive incoming arrows from the others, implying that the song timing of this species may not be affected by the other species' song timings. On the other hand, the NAFL receives two incoming arrows from the BAWF and COTI, implying that this species may decide song timings depending on the other species' timing (e.g., deciding not to sing if others are singing, or to stop singing if others begin to sing). This result appears to fit with intuitive observation because the BAWF sings loudly and consistently on top of a tree whereas the NAFL sings and moves around in the middle of a tree. Although the result is from a short recording, we have observed similar tendencies in some other recordings.

4. Conclusion

We introduced HARKBird and demonstrated that it successfully estimated the degree of arrival and separated sounds of bird songs, with some localization results of songs recorded in two different types of forests in the USA and Japan. We summarized the advantages and limitations of the system. Results showed that HARKBird roughly grasped the soundscape consisting of bird songs of multiple species. If conditions permit, HARKBird localizes over 70% of songs as sound sources. This means that localization results can reduce two costs: the cognitive cost for annotation by knowing the spatial or directional relationships among individuals, and the operating cost by using localization results as a template or initial data of annotation.

The preliminary analysis of bird songs recorded in Inabu, Japan, also showed that HARKBird roughly captures the singing behavior of different species in the soundscape. However, at the same time, the analysis highlighted several technical limitations of HARKBird. It localized many unnecessary sounds that may be associated with abiotic features of the acoustic environment. In addition, we still need to discriminate manually between songs of different species at this phase. These issues can be resolved by more flexible tuning of the parameter of HARK [19] and automatic classification of sound sources.

We also showed the existence of temporal overlap avoidance in the singing behaviors of some species in the recording in Japan, and discussed the asymmetric relationships among species by using transfer entropy analysis. This shows that HARKBird has a high potential to reveal complex interaction processes that underlie temporal soundscape partitioning among species.

We are currently developing a function that automatically recognizes and separates bird songs using with a deep neural network technology. We believe that further development of HARKBird will improve our understanding of such complex behaviors in bird communities. HARKBird is now available with a new inexpensive commercial microphone array, TAMAGO.

Acknowledgements

We thank Charles Taylor and Martin Cody (UCLA) for supporting recordings in the USA; Takashi Kondo and Naoki Takabe (Nagoya University) for supporting field work in Japan; and Greg Gilson and Neil Losin for providing photos of birds in California. This work was supported in part by JSPS KAKENHI: 15K00335, 16K00294 and 24220006.

References:

[1] C. K. Catchpole and P. J. B. Slater, "Bird Song: Biological Themes and Variations," Cambridge University Press, 2008.

[2] M. L. Cody and J. H. Brown, "Song asynchrony in neighbouring bird species," Nature, Vol.222, pp. 778-780, 1969.

[3] R. W. Ficken, M. S. Ficken, and J. P. Hailman, "Temporal pattern shifts to avoid acoustic interference in singing birds," Science, Vol.183, No.4126, pp. 762-763, 1974.

[4] J. W. Popp, R. W. Ficken, and J. A. Reinartz, "Short-term temporal avoidance of interspecific acoustic interference among forest birds," Auk, Vol.102, pp. 744-748, 1985.

[5] R. Planqué and H. Slabbekoorn, "Spectral overlap in songs and temporal avoidance in a peruvian bird assemblage," Ethology, Vol.114, pp. 262-271, 2008.

[6] R. Suzuki, C. E. Taylor, and M. L. Cody, "Soundscape partitioning to increase communication efficiency in bird communities," Artificial Life and Robotics, Vol.17, No.1, pp. 30-34, 2012.

[7] X. Yang, X. Ma, and H. Slabbekoorn, "Timing vocal behaviour: Experimental evidence for song overlap avoidance in Eurasian Wrens," Behavioural Processes, Vol.103, pp. 84-90, 2014.

[8] C. Masco, S. Allesina, D. J. Mennill, and S. Pruett-Jones, "The song overlap null model generator (song): a new tool for distinguish-

7. We used $k = l = 1$.

ing between random and non-random song overlap," Bioacoustics, Vol.25, pp. 29-40, 2016.
[9] R. Suzuki and T. Arita, "Emergence of a dynamic resource partitioning based on the coevolution of phenotypic plasticity in sympatric species," J. of Theoretical Biology, Vol.352, pp. 51-59, 2014.
[10] R. Suzuki and M. L. Cody, "Complex systems approaches to temporal soundspace partitioning in bird communities as a self-organizing phenomenon based on behavioral plasticity," Proc. of the 20th Int. Symposium on Artificial Life and Robotics, pp. 11-15, 2015.
[11] R. Suzuki, R. Hedley, and M. L. Cody, "Exploring temporal soundspace partitioning in bird communities emerging from inter- and intra-specific variations in behavioral plasticity using a microphone array," Abstract Book of the 2015 Joint Meeting of the American Ornithologists' Union and the Cooper Ornithological Society, p. 86, 2015.
[12] I. Aihara, T. Mizumoto, T. Otsuka, H. Awano, K. Nagira, H. G. Okuno, and K. Aihara, "Spatio-temporal dynamics in collective frog choruses examined by mathematical modeling and field observations," Scientific Reports, Vol.4, Article ID: 3891, 2014.
[13] J. Degesys, I. Rose, A. Patel, and R. Nagpal, "DESYNC: Self-organizing desynchronization and TDMA on wireless sensor networks," Int. Conf. on Information Processing in Sensor Networks (IPSN), pp. 11-20, 2007.
[14] 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," J. of Applied Ecology, Vol.48, pp. 758-767, 2011.
[15] T. C. Collier, A. N. G. Kirschel, and C. E. Taylor, "Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network," The J. of the Acoustical Society of America, Vol.128, pp. 182-189, 2010.
[16] Z. Harlow, T. Collier, V. Burkholder, and C. E. Taylor, "Acoustic 3d localization of a tropical songbird," IEEE China Summit and Int. Conf. on Signal and Information Processing (ChinaSIP), pp. 220-224, 2013.
[17] D. J. Mennill, M. Battiston, and D. R. Wilson, "Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour," Methods in Ecology and Evolution, pp. 704-712, 2012.
[18] D. J. Mennill, J. M. Burt, K. M. Fristrup, and S. L. Vehrencamp, "Accuracy of an acoustic location system for monitoring the position of duetting songbirds in tropical forest," The J. of the Acoustical Society of America, Vol.119, No.5, pp. 2832-2839, 2006.
[19] R. Suzuki, S. Matsubayashi, R. Hedley, K. Nakadai, and H. G. Okuno, "Localizing bird songs using an open source robot audition system with a microphone array," Proc. of The 17th Annual Meeting of the Int. Speech Communication Association (INTERSPEECH 2016), pp. 2626-2630, 2016.
[20] K. Nakadai, T. Takahashi, H. G. Okuno, H. Nakajima, Y. Hasegawa, and H. Tsujino, "Design and implementation of robot audition system 'HARK' – open source software for listening to three simultaneous speakers," Advanced Robotics, Vol.24, pp. 739-761, 2010.
[21] D. R. Wilson, M. Battiston, J. Brzustowski, and D. J. Mennill, "Sound Finder: a new software approach for localizing animals recorded with a microphone array," Bioacoustics, Vol.23, No.2, pp. 99-112, 2014.
[22] T. Schreiber, "Measuring information transfer," Physical Review Letters, Vol.85, pp.461-464, 2000.
[23] R. Marschinski and H. Kantz, "Analysing the information flow between financial time series: An improved estimator for transfer entropy," The European Physical J. B, Vol.30, pp. 275-281, 2002.

Name:
Reiji Suzuki

Affiliation:
Associate Professor, Graduate School of Information Science, Nagoya University

Address:
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Brief Biographical History:
2003 Received Ph.D. degree from Nagoya University
2003-2007 Research Associate, Nagoya University
2007-2010 Assistant Professor, Nagoya University
2010-2011 Visiting Scholar, University of California, Los Angeles
2010- Associate Professor, Nagoya University

Main Works:
• R. Suzuki, S. Matsubayashi, K. Nakadai, and H. G. Okuno, "Localizing bird songs using an open source robot audition system with a microphone array," Proc. of The 17th Annual Meeting of the Int. Speech Communication Association (INTERSPEECH 2016), pp. 2626-2630, 2016.

Membership in Academic Societies:
• International Society of Artificial Life (ISAL)
• Information Processing Society of Japan (IPSJ)
• The Japanese Society for Artificial Intelligence (JSAI)
• The Society of Instrument and Control Engineers (SICE)
• Japanese Society for Mathematical Biology (JSMB)
• Society of Evolutionary Study, Japan (SESJ)
• The Ornithological Society of Japan (OSJ)

Name:
Shiho Matsubayashi

Affiliation:
Research Collaborator, Graduate School of Information Science, Nagoya University

Address:
Furo-cho, Chikusa-ku, Nagoya City, Aichi 464-8601, Japan

Brief Biographical History:
2005 Received Dual Master's degrees in Environmental Management and Forestry from Nicholas School of the Environment, Duke University
2013 Received Ph.D. in Environmental Science from University of North Texas
2015- Researcher Collaborator, Graduate School of Information Science, Nagoya University

Membership in Academic Societies:
• The Japanese Society for Artificial Intelligence (JSAI)
• The Ornithological Society of Japan (OSJ)

Name:
Richard W. Hedley

Affiliation:
Postdoctoral Researcher, University of California, Los Angeles

Address:
Room 3113, 621 Charles E. Young Drive South, Los Angeles, California 90066, USA

Brief Biographical History:
2012-2016 Ph.D. Student, University of California, Los Angeles (UCLA)
2016 Received Ph.D. degree from UCLA
2016- Present Postdoctoral Researcher, UCLA

Main Works:
• R. W. Hedley, "Complexity, predictability, and time homogeneity of syntax in the songs of Cassin's Vireo (Vireo cassinii)," PLoS One 11:e0150822, 2016.

Membership in Academic Societies:
• American Ornithologists' Union (AOU)

Name:
Kazuhiro Nakadai

Affiliation:
Honda Research Institute Japan Co., Ltd.
Tokyo Institute of Technology

Address:
8-1 Honcho, Wako-shi, Saitama 351-0188, Japan
2-12-1-W30 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

Brief Biographical History:
1995 Received M.E. from The University of Tokyo
1995-1999 Engineer, Nippon Telegraph and Telephone and NTT Comware
1999-2003 Researcher, Kitano Symbiotic Systems Project, ERATO, JST
2003 Received Ph.D. from The University of Tokyo
2003-2009 Senior Researcher, Honda Research Institute Japan Co., Ltd.
2006-2010 Visiting Associate Professor, Tokyo Institute of Technology
2010- Principal Researcher, Honda Research Institute Japan Co., Ltd.
2011- Visiting Professor, Tokyo Institute of Technology
2011- Visiting Professor, Waseda University

Main Works:
• K. Nakamura et al., "A real-time super-resolution robot audition system that improves the robustness of simultaneous speech recognition," Advanced Robotics, Vol.27, Issue 12, pp. 933-945, 2013 (Received Best Paper Award).
• H. Miura et al., "SLAM-based Online Calibration for Asynchronous Microphone Array," Advanced Robotics, Vol.26, No.17, pp. 1941-1965, 2012.
• R. Takeda et al., "Efficient Blind Dereverberation and Echo Cancellation based on Independent Component Analysis for Actual Acoustic Signals," Neural Computation, Vol.24, No.1, pp. 234-272, 2012.
• K. Nakadai et al., "Design and Implementation of Robot Audition System "HARK"," Advanced Robotics, Vol.24, No.5-6, pp. 739-761, 2010.
• K. Nakadai et al., Speech Communication, Vol.44, pp. 97-112, 2004.

Membership in Academic Societies:
• The Robotics Society of Japan (RSJ)
• The Japanese Society for Artificial Intelligence (JSAI)
• The Acoustic Society of Japan (ASJ)
• Information Processing Society of Japan (IPSJ)
• Human Interface Society (HIS)
• International Speech and Communication Association (ISCA)
• The Institute of Electrical and Electronics Engineers (IEEE)

Name:
Hiroshi G. Okuno

Affiliation:
Professor, Graduate School of Science and Engineering, Waseda University
Professor Emeritus, Kyoto University

Address:
Lambdax Bldg 3F, 2-4-12 Okubo, Shinjuku, Tokyo 169-0072, Japan

Brief Biographical History:
1996 Received Ph.D. of Engineering from Graduate School of Engineering, The University of Tokyo
2001-2014 Professor, Graduate School of Informatics, Kyoto University
2014- Professor, Graduate School of Science and Engineering, Waseda University

Main Works:
• "Design and Implementation of Robot Audition System "HARK"," Advanced Robotics, Vol.24, No.5-6, pp. 739-761, 2010.
• "Computational Auditory Scene Analysis," Lawrence Erlbaum Associates, Mahmoh, NJ, 1998.

Membership in Academic Societies:
• The Institute of Electrical and Electronic Engineers (IEEE), Fellow
• The Japanese Society for Artificial Intelligence (JSAI), Fellow
• Information Processing Society Japan (IPSJ), Fellow
• The Robotics Society of Japan (RSJ), Fellow