Paper:

Acoustic Monitoring of the Great Reed Warbler Using Multiple Microphone Arrays and Robot Audition

Shiho Matsubayashi[*1], Reiji Suzuki[*1], Fumiyuki Saito[*2], Tatsuyoshi Murate[*2], Tomohisa Masuda[*2], Koichi Yamamoto[*2], Ryosuke Kojima[*3], Kazuhiro Nakadai[*4,*5], and Hiroshi G. Okuno[*6]

[*1]Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
E-mail: mt.shiho@gmail.com, reiji@nagoya-u.jp
[*2]IDEA Consultants, Inc., Japan
E-mail: {fumiyuki, murate, msd20750, ykouichi}@ideacon.co.jp
[*3]Graduate School of Information Science and Engineering, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
E-mail: kojima@cyb.mei.titech.ac.jp
[*4]Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
[*5]Honda Research Institute Japan Co., Ltd.
8-1 Honcho, Wako-shi, Saitama 351-0188, Japan
E-mail: nakadai@jp.honda-ri.com
[*6]Graduate School of Fundamental Science and Engineering, Waseda University
2-4-12 Okubo, Shinjuku, Tokyo 169-0072, Japan
E-mail: okuno@aoni.waseda.jp
[Received July 30, 2016; accepted November 1, 2016]

This paper reports the results of our field test of HARKBird, a portable system that consists of robot audition, a laptop PC, and omnidirectional microphone arrays. We assessed its localization accuracy to monitor songs of the great reed warbler (*Acrocephalus arundinaceus*) in time and two-dimensional space by comparing locational and temporal data collected by human observers and HARKBird. Our analysis revealed that stationarity of the singing individual affected the spatial accuracy. Temporally, HARKBird successfully captured the exact song duration in seconds, which cannot be easily achieved by human observers. The data derived from HARKBird suggest that one of the warbler males dominated the sound space. Given the assumption that the cost of the singing activity is represented by song duration in relation to the total recording session, this particular male paid a higher cost of singing, possibly to win the territory of best quality. Overall, this study demonstrated the high potential of HARKBird as an effective alternative to the point count method to survey bird songs in the field.

Keywords: acoustic monitoring, microphone arrays, robot audition, HARKBird, the great reed warbler

1. Introduction

The recent advances in acoustic engineering, microphone arrays in particular, offer an emerging approach for wildlife researchers to acoustically monitor species of interest for a long period without human interruptions [1]. Ornithologists, in particular, will greatly benefit from this technology to passively monitor the movements and behaviors of birds through their songs.

One of the most advantageous features of microphone arrays over conventional single microphones when monitoring birds is its ability to detect the direction of arrival (DOA) of the sound event. Using the DOA of sound events acquired from multiple microphone arrays, we can determine the position of the sound source in two-dimensional (2D) space. Localization using microphone arrays has been proven effective to auditorily distinguish different species or individuals singing simultaneously [2, 3] and track movements of individuals [4], both of which can not be easily achieved by conventional microphones. Microphone arrays, therefore, have a great potential for field ecologists to capture auditory scenes where individual birds acoustically interact with other members of the community. By analyzing the acoustic interactions of birds derived from microphone arrays, we can possibly reveal hierarchical competitions for limited resources, e.g., habitat space or sound space.

Despite the potential for microphone arrays to passively observe birds for a long period, microphone arrays

have not been widely used in bird surveys mainly owing to difficulties obtaining the equipment and implementing the system in the field [5]. To overcome these challenges, we developed HARKBird,[1] a portable system that consists of a laptop computer, open source robot audition system HARK (Honda research institute, Audition for Robot with Kyoto university)[2] [6], and commercially available low-cost microphone arrays. HARKBird provides a GUI to record, localize,[3] separate, and export outputs for annotation using the HARK network. The entire software system is composed of a series of Python scripts with modules (e.g., wxpython, pyside) and other standard sound processing software (e.g., sox, arecord, aplay) that operate under Ubuntu Linux 12.04, where the latest HARK and HARK-Python are installed [2].

The HARK sound source localization algorithm is based on the MUltiple SIgnal Classification (MUSIC) method [7] using multiple spectrograms with the short-time Fourier transformation (STFT). Localized sounds are then separated to multiple songs using the Geometric High-order Decorrelation-based Sound Separation (GHDSS) method [8] in real time. HARKBird allows users to adjust three parameters to optimize localization performance: the expected number of sound sources to determine the number of sound sources in the track, the lower bound frequency for MUSIC to reduce noises in localization, and the threshold for source tracking to control noise [2].

Using HARKBird and prerecorded bird songs in a conifer-mixed forest in Japan, we confirmed a sufficient level of accuracy to estimate the DOA of 11 different bird's songs played through a speaker [9]. The pilot studies demonstrated HARKBird's potential to grasp the acoustical interaction of birds in detail [2, 9]; however, the localization accuracy of actual birds in the field was not fully explored.

The objective of this study is to assess the localization accuracy of birds in 2D space and time using HARKBird and three omnidirectional microphone arrays. We chose to record the great reed warbler because localization results of their loud songs from song posts in a reed marsh can be approximated to 2D space. Note that as a future work, after building up a technical foundation, we plan to extend the use of HARKBird to three-dimensional space. To achieve our objective, we compared the localization results with the actual location of the birds and duration of vocalization detected by trained human observers. Specifically, we conducted bird recordings and observations in a relatively open reed marsh where visual obstacles were minimal. A clear view of the target species in a field minimizes potential observer bias associated with the position of the birds. It also permits observers to easily identify the timing of a song, i.e., the beginning and ending of vocalization. Finally, recording experiments in a reed marsh with sparsely distributed trees allow sound distortion between the bird and microphone arrays to be minimized, hence increasing localization performance.

2. Methods

2.1. Target Species

Our target species, the great reed warbler (*Acrocephalus arundinaceus*), is a polygamous passerine that inhabits a reed marsh. Males of the great reed warbler actively defend their territories by loud and persistent songs perched on their song posts. Songs of the great reed warbler have long been studied to understand the mating selection of females. Several field studies suggested that females of the great reed warbler produce a larger clutch size by mating with males who have larger repertoire [10, 11]. Conversely, a recent study indicated that their analysis overlooked the importance of several confounding factors such as male age, differential attraction, and variation in habitat quality, concluding that territory quality was a stronger predictor of harem size than male repertoire size [12]. That is, a male that defends a habitat of higher quality may have higher reproductive success. In either case, male great reed warblers sing to attract females, directly or indirectly, and to discourage other neighboring males of the same species that could possibly invade his territory; thus, it is a useful indicator to measure his dominance.

Songs of the great reed warbler vary in type and length. For example, **Fig. 1** presents a spectrogram of two male great reed warbler vocalizations, "a" and "b" in our recording. When a male sings, he persistently repeats songs separated by short intervals. This assemblage of songs and in-between intervals frequently lasts from a few to over ten minutes. The song of "b," referred to as "b1," appears faint compared to the songs of "a" because he was singing more distant from the microphone arrays. As explained in detail later, HARK localized and separated each song in the majority of the cases, with exceptions of occasional mislocalization. In the case of mislocalization, HARK generated a long vocalization, which in fact was the songs of multiple individuals.

2.2. Bird Observation

We conducted a bird survey in mid-May, during the early breeding season of 2016, on a bank of the Ibi river, Kaminogo district, Mie prefecture in central Japan (35°34′59″, 136°06′29″). Using spotting scopes, observers recorded the identification of the bird based on color bands tagged to both legs (**Fig. 2**), location, activities, and timing of each activity of the bird. Because human observers cannot record the exact beginning and ending timings of each song (see details for **Fig. 1**), we instead reported the beginning and the ending timings of a series of songs including short breaks in-between. Hereafter, this time period is referred to as the directly observed song duration.

1. Available from our website: http://www.alife.cs.is.nagoya-u.ac.jp/~reiji/HARKBird/ [Accessed January 20, 2017]
2. Retrieved July 6, 2016 from http://www.hark.jp/document/hark-document-en/
3. We refer to the term localization as estimating the DOA of the sound event.

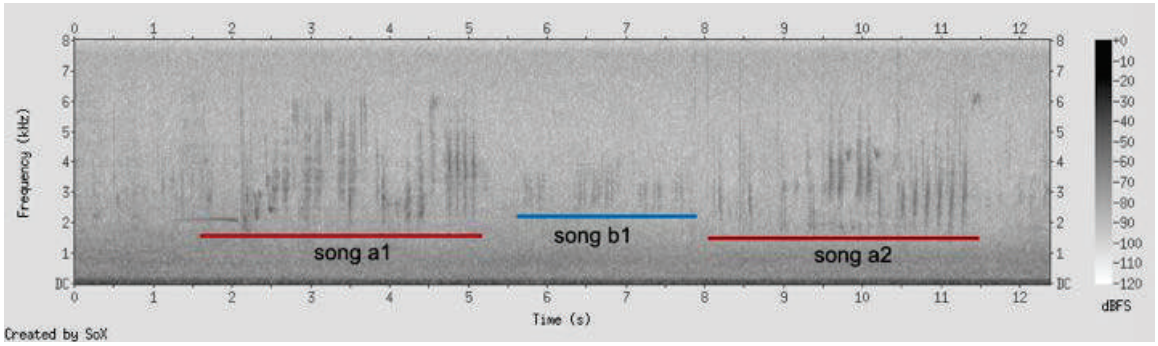

Fig. 1. Sample spectrogram of the great reed warbler's vocalization in our recording. Each solid line indicates the duration of a single song of two males, "a" and "b." Note songs referred to as "a1" and "a2" are different songs of "a."

Fig. 2. Banded male great reed warbler. This one wears leg bands in three different colors, red, yellow, and brown from top to bottom.

Fig. 3. Photo illustrating one of the three microphone arrays connected to a laptop computer at the recording site.

Bird observation was conducted for approximately 6 h, starting at 6:00 AM. Each observation session was 20 min long, which corresponded to the recording session that occurred at the same time.[4] Observers determined the location of each bird using landmarks on drone imagery and marked posts in the field. The positions of the landmarks were measured using a GPS.[5]

Within the survey area, three warbler males were banded for identification prior to this recording experiment, two of whom were visually confirmed during recording sessions. While recording, we confirmed the presence of a maximum of five individuals at one point in our experiment, two of whom were banded males. According to the field observation, there were a small number of additional males in the study area, at least one of whom was visually confirmed to be a male without bands. This unbanded male briefly flew in and out of the study area, which represents typical behavior of a young male floater in search of vacant territory.

2.3. Placement of Microphone Arrays

We used microphone arrays called Tamago,[6] an egg-shaped body, 8 cm in diameter and 12 cm in height. Tamago has eight microphones arranged every 45° horizontally along the middle circumference of the body. We placed each microphone array 1.2 m above the ground using a tripod (**Fig. 3**). We linearly placed three microphone arrays in a row, each of which was approximately 15 m apart, on the riverbank. We chose to record on the riverbank where the sound transference was maximized such that we could record the target species over a wide range. The linear spatial arrangement was most efficient to connect microphone arrays to a laptop PC[7] using USB cables.

2.4. 2D Localization and Estimating Song Duration Using HARKBird

2.4.1. HARKBird

We synchronized recording and estimated the DOA of the sound source acquired from each microphone array

4. We had a 10 min break between each recording session.
5. Trimble R10, GNSS.
6. TAMAGO-01 by System in Frontier, Ltd.
7. Panasonic TOUGHBOOK CF-C2.

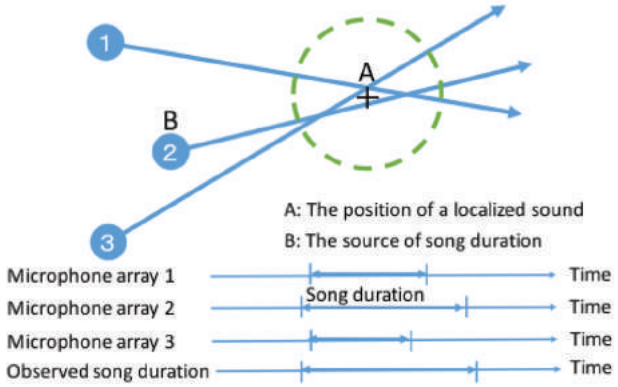

Fig. 4. Diagram illustrating a conceptual model to obtain (A) the position of a localized song, and (B) song duration.

using HARKBird. We extended the original scripts of HARKBird such that all three microphone arrays connected to a laptop could start and stop at the same time. We applied the HARK network embedded in HARKBird for each recorded track to localize and separate the recorded wave files in the following steps. After reading the sound signals obtained from the eight channels for each microphone array, the network converted these signals to 16 Hz. It then localized and estimated the DOA of the sound sources in the spectrograms using the MUSIC method with the STFT. Localized sounds were then separated into multiple songs using the GHDSS method for the recorded track acquired from each microphone array.

Finally, we integrated the localized sounds from the three microphone arrays to extract the directional and temporal information of the sound source as described below.

2.4.2. Spatial Data

At each timeframe of the localization process, we assumed that a half line arose from each microphone array towards the localized sound source, i.e., the song of a great reed warbler as illustrated in **Fig. 4**. We used the center of mass of the three intersections of those three half lines as the estimated location of the localized sound (A in **Fig. 4**) when all the intersections were contained within the range of 30 m (indicated as a dotted circle in **Fig. 4**) from the center of mass. Because each sound was localized during multiple time frames, we adopted the mean position of the sound as its location. Note that we assumed the stationarity of the bird for simplicity, i.e., the bird remained at one particular location from the beginning to end of a song.

2.4.3. Temporal Data

The localization performance of the microphone arrays decreases significantly with increasing distance from the sound source. If the sound was not fully detected by all three microphone arrays, it was rejected as a false detection. Even when all three microphone arrays partially detected fragmented sounds at a distance, the song duration could be shorter than the actual songs, resulting in an underestimation of the song duration.

We overcame this issue by adopting the song duration localized at the closest microphone array to the sound source. More specifically, we employed the duration of a song localized by the microphone array that was the closest to the center of gravity of the triangle discussed in the previous section. For example, in **Fig. 4**, we used the song duration localized by microphone array 2, rather than the song duration localized by the three microphone arrays at the center of the gravity of the three half lines, A. We thus relied on the closest microphone array to the bird when obtaining the song duration.

2.5. Accuracy Assessment

Assuming a high correlation between the position of the localized songs and the actual location of the singing males, we classified the localized sounds as songs of the observed bird based on the distance. Considering the positional misalignment of localized sounds acquired from three microphone arrays, we considered all localized sounds within a 20 m radius circle around each of the singing male as songs of the individual at the center. We first spatially extracted all the separated sounds contained within this search distance of 20 m. We then manually classified these localized sounds, presumably the great reed warbler songs according to the timing of each observed duration, e.g., the beginning and ending of a series of songs, recorded by the human observers. More specifically, for each individual, we assembled the duration of a series of localized songs to the song duration if the song duration was contained within the directly observed song duration of the corresponding individual. Owing to the difficulties of recording the exact timing of a song in units of seconds, in addition to the potential time lag between observers' watches, we provided each song a margin of 30 s before and after. Because the song posts of the two-banded males were approximately 70 m away, the former step separated their songs, except for several individuals singing simultaneously in close proximity. Further, the latter step reduced the risk of double counting localized songs of one male as those of multiple males, and vice versa, when they were singing alternately in close proximity.

We further assumed that the positional and temporal data collected by the human observers were correct and used this dataset as a baseline to assess localization accuracy in both space and time. We compared the localized dataset, namely, bird identity, position, and duration of each song, all of which were generated by the two steps described above, with the dataset derived from the direct human observation. More specifically, we assessed the positional accuracy by calculating the root-mean-square error (RMSE), which measures the positional difference using *x*-*y* coordinates of the localized sounds and that of the actual birds. We next assessed the temporal accuracy by comparing the total song duration, which included all songs localized by HARKBird for each bird and the di-

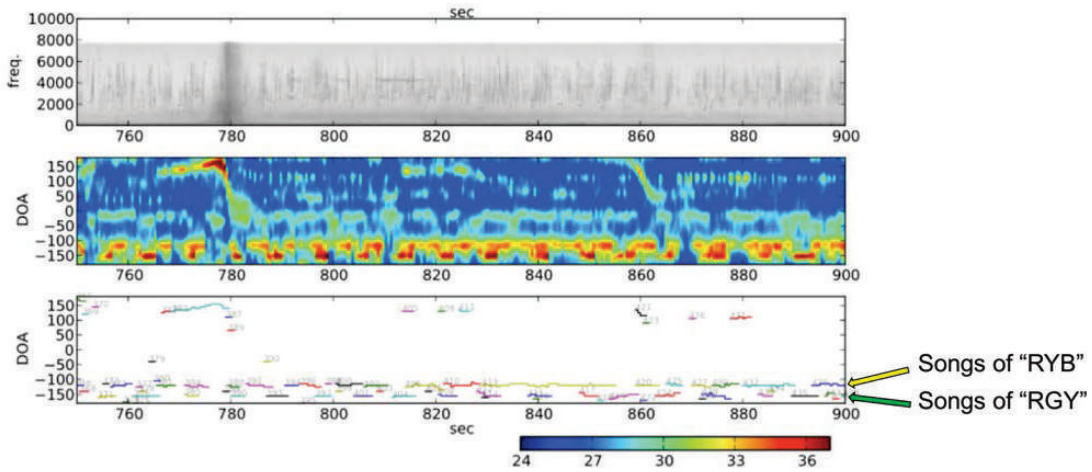

Fig. 5. 2.5 min duration auditory scene captured by HARKBird: the spectrogram of a channel of the original recording (top panel), the MUSIC spectrum (middle panel), and the DOA and timing of each separated sound (bottom panel) derived from one of the three microphone arrays.

rectly observed song duration recorded by the human observers based on the beginning and ending timing, for each recording session.

3. Results

3.1. Parameter Settings

We set three HARKBird parameters to analyze the recorded sounds. First, we set the number of sound sources, which enhances the peak of the target sound, to three. To maximize the localization performance of the great reed warbler, which projects songs loudly from its song post, we chose a relatively high threshold value of MUSIC spectrum, 31.5. We also set the lower-bound frequency to 2200 Hz to capture the lower limit of the great reed warbler. See [2] for details of the parameter settings.

3.2. Spatial Components

We visually or auditorily confirmed the presence of a maximum four individuals during four consecutive recording sessions, starting at 11:00 AM. Of the four detected individuals, we visually confirmed that two were wearing the bands we tagged. The other two, classified as "Unknown," were confirmed only by hearing. Thus they could be either a third and/or fourth male, or the two banded males that disappeared from the observer's sight for a moment. Based on the field observation, i.e., the position of each bird and the timing of the songs, the songs of the "Unknown" in sessions 11 and 12 could be the songs of the banded male, "RYB." The songs of the "Unknown" in sessions 13 and 14 were likely to be the songs of another male based on the field observations, timing of the songs, and territory defending behaviors other than territorial songs a banded male displayed during the sessions.

Figure 5 illustrates an auditory scene captured by one of the three microphone arrays for a recording session starting at 11:30 AM. Based on the DOAs and field observation, we estimated the presence of two singing males, "RYB" and "RGY" at approximately -120° and -150°, respectively, from the microphone array.

Figure 6 displays the spatial distribution pattern of the birds recorded by the human observers and localized sounds during four consecutive recording sessions. As can be observed, there were possibly three males, "RYB," "RGY," and "Unknown" on the scene. Unlike the former two, the appearance of the "Unknown" individual was not visually confirmed by the human observers. The songs of the "Unknown," therefore, could be the songs of other males including the banded males.

Notably, numerous noises appeared across the experimental site (**Fig. 6**), the majority of which were less than a second in duration. These noises reflected a high occurrence of false detection of sound events. False detection can be caused by various reasons, for example, other bird species temporally singing in the area; a Japanese skylark (*Alauda Japonica*), could have triggered accidental localization. In this case, we delineated the warbler songs from the other birds' songs by considering the differences in behavioral characteristics: the great reed warbler stayed at one song post while they sang, whereas other birds sang and moved around constantly while they foraged. More specifically, we counted how frequently sounds were localized within a small area, a square of 16 m^2 (4 m $\times$ 4 m), such that frequently localized sounds within this small area were highly likely to be the songs of the great reed warbler given its high site fidelity.

The other type of accidental mislocalization occurred when each microphone array localized different individuals instead of one. For example, multiple songs can overlap each other when multiple males sing simultaneously or alternately. In either manner, the microphone arrays could respond to multiple singers differently, depending on the relative position of the sound sources and microphone array. We found, however, that the solution to the first issue, i.e., only adopting sounds localized at a high

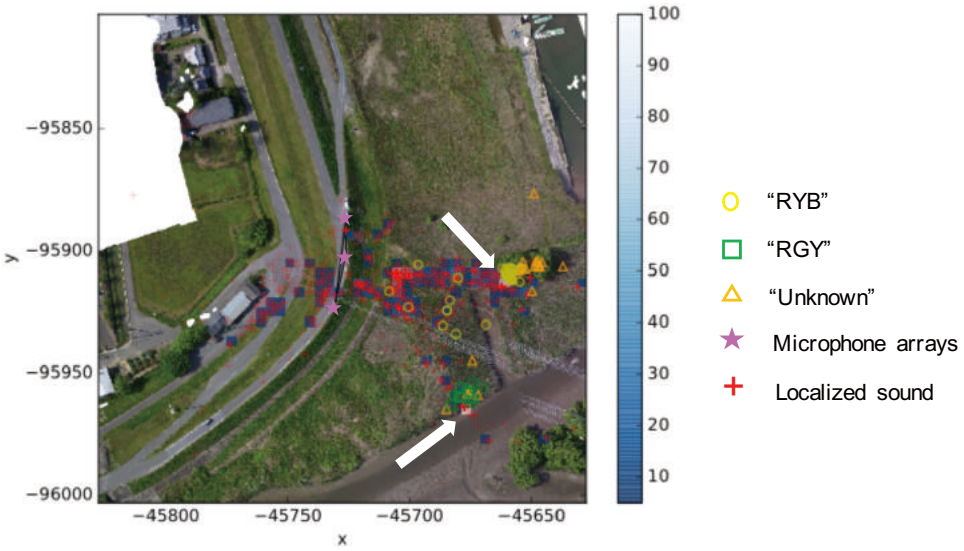

Fig. 6. Spatial distribution pattern of observed birds and localized sounds. Three microphone arrays, each of the localized sound, and position of singing males detected by human observers are displayed accordingly. Filled squares indicate the frequency of localization within a small area (a square of 4 m × 4 m), with the white being the highest frequency. White arrows identify the positions of the two singing males that were localized most frequently.

Table 1. RMSE values calculated for possible three males. Each value is the average of all localized songs of the corresponding individual.

	RYB	RGY	UNK
Session 11	18.53	19.30	39.63
Session 12	21.87	14.70	19.48
Session 13	24.18	9.87	14.37
Session 14	26.85	13.00	13.65
Average	22.86	14.22	21.78

frequency within a small area, was proven effective to diminish the influence of this type of accidental localization (**Fig. 6**). This may suggest that males of the great reed warbler in fact sing alternately to avoid overlap.

We determined that two spots within the study area were constantly localized with a high frequency, after removing those accidentally localized sounds (**Fig. 6**). Clearly, these two spots corresponded to the position of the two banded males identified by the human observers, with the northern spot representing “RYB” singing in a tree (indicated by solid line circles in **Fig. 6**) and the southern spot representing “RGY” singing at the waterfront (indicated by solid line squares in **Fig. 6**).

We next assessed the positional accuracy of the localization using the root-mean-square error (RMSE). We calculated the RMSE to measure the positional difference between the localized bird songs and that of the actual birds derived from the directly observed song duration. For simplicity, we only focused on the two banded males coded as “RYB” and “RGY.”

As summarized in **Table 1**, the RMSE values of “RGY” were primarily less than the RMSE values of the other males throughout four consecutive recording sessions, indicating the highest positional accuracy. The lowest average RMSE value of “RGY” likely reflects the stationarity of this individual that sang at one particular song post throughout the experiment. Conversely, other males were more mobile. In particular, “RYB,” the other banded male, was highly active in session 14, demonstrating an obvious territory defending behavior by flying at the periphery of his territory (**Fig. 6**). Similarly, “Unknown,” which was likely to be a young male floater in search for a territory, demonstrated higher RMSE values than the “RGY.” Alternatively, the higher RMSE values of “RYB” and “Unknown,” other than session 14 when the territory patrolling behavior was clear, may suggest the misclassifying of the songs of multiple males (see **Fig. 1** for details) in close proximity.

3.3. Temporal Components

We first compared the total song duration localized by HARKBird and the directly observed song duration by the human observers.[8] As displayed in **Fig. 7**, the total song duration localized by HARKBird was consistently 30%–50% shorter than the directly observed song duration throughout four sessions. Shorter song duration could be explained by the difference between how a song was perceived by robot audition and human observers.

More specifically, HARKBird only recognized the vocal part of a song in the form of a series of syllables, e.g., a1, a2, in **Fig. 1**. Conversely, human observers included both the vocal part and silent window period in-between when reporting the directly observed song duration. That

8. We found that HARKBird tended to add a short blank time of several milli seconds, when it localized a sound event. We did not consider this blank time for further analysis assuming that the effect was marginal.

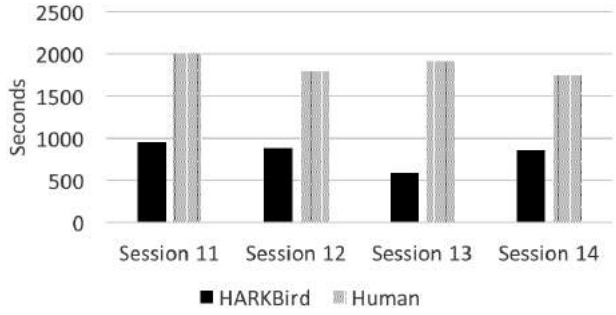

Fig. 7. Total song duration localized by HARKBird and human observers.

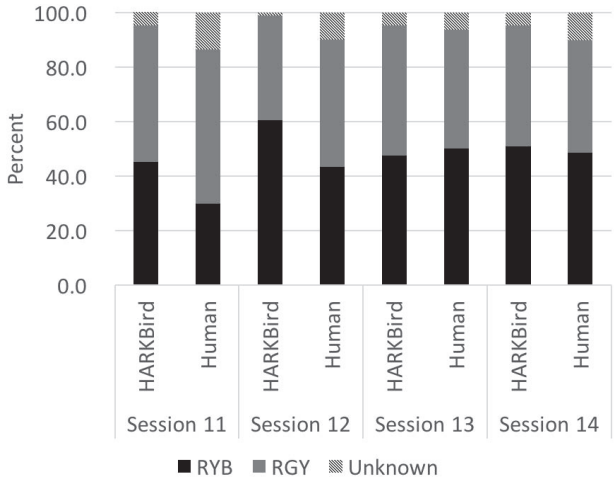

Fig. 8. Percent of total duration by each male measured by HARKBird and human observers.

is, the silent part was processed as a part of the observed duration only by human ears. Further, note that the directly observed song duration only referred to the beginning and ending of vocalization, which include songs and the window period in various lengths. Moreover, we cannot deny the possible mislocalization of songs, especially when the great reed warbler was singing farther than over the distance limit of the microphone arrays to localize sounds.

We next examined whether the localized song duration and the directly observed duration had a similar ratio among each bird in terms of a percent of the total song duration. **Fig. 8** displays the percent of total duration by each male measured by HARKBird and the human observers. Clearly, the banded males "RYB" and "RGY" had a higher percent to the total than "Unknown" throughout the four sessions. Further, both methods generated a similar ratio in two of the four sessions.[9] For example, in session 14, the localized song duration revealed that "RYB" had the highest percent (50.8%), followed by "RGY"(44.6%), which was consistent with the highest percent of "RYB"(48.6%) and "RGY"(41.1%) indicated by the total observed song duration. Conversely, the ratio calculated for the localized and observed song duration did not match in the remaining sessions.[10] For example, in session 13, the localized song duration presented a marginally higher percent for "RGY" (47.7%) than "RYB" (47.4%), whereas the observed data generated the opposite result, with 43.7% and 50.0% for "RGY" and "RYB," respectively.

Three technical difficulties could be the reasons for this mismatched ratio between the localized and observed data. First, the exact identification of the singer was challenging, especially when the observers only had audio cues. The songs of "Unknown" could have belonged to the two-banded individuals that temporarily disappeared from the observer's sight, or other individual(s), possibly young male floaters that were attempting to invade the territories of the two-banded males. Secondly, the relative position of the microphone arrays and singing males could have affected the localization performance. If multiple individuals were singing in close proximity, or in a similar direction, it was difficult to distinguish the two based on the DOA. Finally, as discussed, HARK could have failed to separate the songs and consequently generated a long song for one male warbler, that in fact consisted of the songs of multiple males (see **Fig. 1** for details).

If the dominance of a male was simply characterized by the song duration measured by HARKBird, that is, the vocal part of a song excluding the blank time between each song, one of the singing males, coded as "RYB," which had a longer song duration in two of the four recording sessions,[11] had a dominance over the other banded male, "RGY." Although "RGY" marked a longer song duration in one of the four sessions,[12] the field notes revealed that songs of an unbanded male coded as "Unknown" in that particular session were likely to be the songs of "RYB" that went out of the observer' sight for a period of time. If we recalculated the songs of each bird by adding songs of "Unknown" to that of "RYB," the percent of each bird to the total song duration became 49.9% for "RYB" and 50.1% for "RGY," indicating a marginal difference between the two males in this session. Similarly, in session 13, if the songs of "Unknown"[13] were in fact the songs of the banded male nearby, "RYB" had a longer song duration than "RGY." We can therefore conclude that "RYB" demonstrated a higher dominance in the soundscape.

4. Discussion

4.1. Merits of Using HARKBird to Measure Song Duration

The advantage of using HARKBird in monitoring birds over human observers is its ability to capture the vocal part of each song in units of seconds. This measure-

9. Sessions 11 and 14.

10. Sessions 12 and 13.
11. Sessions 12 and 14.
12. Session 11.
13. There were two "Unknown" warblers recorded in the field. Of the two, one was observed near "RYB" and the other was observed near "RGY."

ment was considerably more accurate than the directly observed song duration by the human observers, which was essentially the beginning and ending timing of an assemblage of songs in units of minutes. This assemblage of songs contained blank intervals of various lengths. That is, the directly observed song duration inherently included both vocal and silent parts. Conversely, HARKBird localized only the vocal part of a song in units of seconds. Furthermore, we noticed that determining the end of a song based only on audio cues was considerably more challenging than determining the beginning of a song, i.e., it was easier for human ears to hear sounds than silence. Difficulties multiplied when multiple individuals were singing simultaneously. Even when we had only one male target, he could sing sporadically over a long period. In either case, subjectivity could be introduced when estimating the end of a vocalization, which was frequently recognized after the male actually stopped singing. HARKBird was particularly useful in reducing those potential sampling biases.

4.2. Behavioral Implications

The rich ecological data obtained from HARKBird creates a new path for ornithologists to better understand the behavioral aspect of bird songs in the soundscape, which cannot be easily achieved by conventional direct point count by human observers. For example, we can deduce the cost of territorial defense by songs from the exact song duration measured by HARKBird. Localization results revealed that one of the singing males, "RYB," sang for an average of 6.8 min during each recording session of 20 min. That is, if the cost of singing was simply measured by the song duration, 33.4% of all energy was consumed by singing. Similarly, the other banded males, "RGY" sang for an average of 6.2 min or 30.9% of energy was consumed by singing. In contrast to these banded individuals, the "Unknown" allocated only 34.5 s, which was equivalent to 2.7% of all energy to singing.[14] This dominance order measured by the percent of song duration by each male to the total recording duration was consistent with the hierarchy measured by the percent of total duration by each male to the total song duration (**Fig. 8**).

For any individual, the amount of resource, e.g., the total amount of energy that one can allocate to various activities, within a particular environment is limited. In the case of the great reed warblers we observed, the remaining energy other than singing could have been used for other activities such as foraging, parental care, territory defense, recruiting vacant territory space, and searching for the chance of extra pair copulation. The higher cost "RYB" paid on singing in exchange for those activities may lead to a higher reproductive success, because he not only could better broadcast his presence to surrounding females but also declare the exclusive use of the territorial space and females within against neighboring rivals.

14. Assuming the song belongs to one individual. We only report the song duration of "Unknown" in sessions 13 and 14 because "Unknown" is highly likely to be the third male in the scene. See 3.2. for details.

Moreover, capturing the exact timing of songs with positional information allows ornithologists to investigate the communication efficiency of birds. A field study in a conifer-oak mixed forest in California has indicated that songbirds may compete for time during which they can maximize the chance to broadcast their territories, in addition to territorial space itself. What this suggests is that birds may divide up the soundscape in a manner that they avoid overlap on purpose, rather than sing at random [13]. Extracting the exact timing of songs with positional information for every member of the scene could be a powerful tool to analyze the auditory scene where multiple individuals compete for soundscape.

5. Limitation

This study contains one major technical limitation, the lack of automatic sound classification. Subjectivity could be reduced if we could distinguish bird songs based purely on the characteristics of the song, such as the structure of a song, without supplementary information, i.e., the position and the approximate timing of the song collected by human observers. Further, when distinguishing multiple individuals that were singing simultaneously, we found that HARKBird's performance to localize birds could have been affected by the relative position of the singing males. Independence from the locational information could increase robustness to the relative position of the singer.

Independence from locational information can be achieved in three manners, namely, visual inspection of spectrograms, semi-automatic classification, and automatic classification of bird songs [14–22]. The visual inspection of the spectrogram is apparently extremely costly and only effective for the analysis of songs in a short period. The requirement for improving efficiency in data analysis has led to the development of the latter two classification methods. Semi-automated classification, where a trained machine classifies bird songs, has proven particularly useful when a human supervisor clearly understands the songs of the target species [14]. More recently, increasing numbers of studies have demonstrated the potential of the the third method, automatic sound classification using machine learning [15–22]. Conceptually, machine learning is advantageous over semi-automated classification in decreasing potential bias created by human supervisors when training a machine, especially either when the singing male has many song variations or sings in noisy environments [22].

Independence from locational information has several significant implications for this study. Most importantly, it reduces the ambiguity associated with "Unknown" individual whose appearance is not visually confirmed in the field. Further, it could increase the robustness to the relative position of the singing individuals. In this study, bird songs were classified based on the spatial position and approximate timing of each male identified by human observers. This process cannot be free from the possibil-

ity of double counting the song duration when we failed to delineate songs of simultaneously singing males in close proximity. More specifically, the songs of "RYB" and "Unknown" could be intermixed and misclassified when they were close by. This scenario could occur with a high probability when neighboring males compete for territories by singing. Moreover, if we could delineate the songs of multiple individuals based on the characteristics of the songs, we could eliminate issues of long and unseparated vocalization, which could be generated because of mislocalization (see **Fig. 1** for details). Further, if we could delineate the songs of individuals without the reference data collected by human observers, we could eliminate potential observer bias inherent to the field data collection both in time and space.

Furthermore, independence from locational information could also be useful to investigate songs at the individual level. The current system was effective for detecting the DOA of each song, yet was limited for distinguishing different song types. Based on our field observation, the great reed warbler males vocalized different types of songs for different purposes: their warning songs sounded shorter and more simply structured compared to the other type of song, the long and persistent one they vocalize on their song posts to broadcast their dominance. Further, another recent behavioral study has indicated that the dominance of the great reed warbler is characterized by the frequency of the switch between the longer and the shorter syllables, rather than the repertoire size which is often considered as a key to measure the dominance in many other passerines [12]. In either case, we must manually investigate the spectrograms for further analysis, which requires a tremendous amount of work. Song classification based purely on the characteristics of the song structure not only increases the efficiency in data analysis but also allows us to study larger datasets, instead of a snapshot of the auditory scene obtained from the manual inspection of spectrograms.

6. Conclusion

HARKBird demonstrated a high potential as an effective alternative to the point count method to survey birds in space and time. Current interests in the use of microphone arrays to monitor birds are motivated by a practical requirement to collect bird songs for a long period without potential observer bias even at a place where visibility is limited. By assessing positional and temporal accuracies to detect songs of the great reed warbler in the field, we confirmed a high applicability of the proposed system to achieve the above requirements. In addition to the accurate measurement of song duration, HARKBird successfully distinguished songs of multiple individuals vocalizing simultaneously.

Rich ecological data such as that demonstrated in this study will significantly benefit many field ecologists and further advance our knowledge of how birds acoustically interact with other members of the community in space and time. Variations in song duration of the great read warblers across recording sessions could be attributed to the behavioral response of each bird, which heavily depend on a variety of random factors in the field and technical challenges in localizing birds in close proximity. For wider applications of this system, future work should focus on the development of a superior data analysis method that increases the robustness to the relative position of the birds under challenging conditions. Moreover, the development of an automatic sound classification method would considerably increase the efficiency in data analysis.

Acknowledgements

We would like to thank Mr. Shinji Sumitani at Nagoya University for assisting in the recording experiment. This work was supported by JSPS KAKENHI 15K00335, 16K00294, and 24220006.

References:

[1] D. Blumstein et al., "Asoustic monitoring in terrestrial environments usnig microphone arrays: applications, technological considerations and prospectus," J. of Applied Ecology, Vol.48, No.3, pp. 758-767, 2011.

[2] 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 2016 Int. Conf. on Spoken Language Processing, San Francisco, Sep 8-12, 2016. doi: 10.21437/Interspeech.2016-782.

[3] T. C. Collier, A. N. G. Kirschel, and C. E. Taylor, "Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network," J. of Acoutical Society of Ameirca, Vol.128, No.1, pp. 182-189, 2010.

[4] A. N. G. Kirschel, M. L. Cody, Z. Harlow, V. J. Promponas, E. E. Vallejo, and C. E. Taylor, "Territorial dynamics of Mexican Antthrushes Formicarius moniliger revealed by individual recognition of their songs," Ibis, Vol.153, pp. 255-268, 2011.

[5] D. J. Mennill, M. Battiston, D. R. Wilson, J. R. Foote, and S. M. Doucet, "Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour," Methods in Eoclogy and Evolution, Vol.3, pp. 704-712, 2012.

[6] 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.

[7] R. O. Schmidt, "Multiple emitter location and and signal parameter estimation on antennas and propagation," IEEE Trans. on Antennas and Propagation, Vol.34, No.3, pp. 276-280, 1986.

[8] H. Nakajima, K. Nakadai, Y. Hasegawa, and H. Tsujino, "Adaptive step-size parameter control for real world blind source separation," Proc. of ICASSP, pp. 149-152, 2008.

[9] S. Matsubayashi, R. Suzuki, R. Kojima, and K. Nakadai, "Hukusu no microphone array to robot chokaku HARK wo mochiita yacyo no ichiseido no kento" (Assessing the accuracy of bird localization derived from multiple microphone arrays and robot audition HARK), Japanese Society of Artificial Intelligence, JSAI Technical Report, SIG-Chellenge-043-11, pp. 54-59, 2015.

[10] C. K. Catchpole, "Song repertoires and reproductive success in the great reed warbler Acrocephalus arundinaceus," Behavioral Ecology and Sociobiology, Vol.19, pp. 439-445, 1986.

[11] D. Hasselquist, S. Bensch, and T. von Schantz, "Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler," Nature, Vol.381, pp. 229-232, 1996.

[12] W. Forstmeyer and B. Leisler, "Repertoire size, sexual selection, and offspring viability in the great reed warbler: changing patterns in space and time," Behavioral Ecology, Vol.15, No.4, pp. 555-563, 2004.

[13] R. Suzuki, C. E. Taylor, and M. L. Cody, "Soundscape partitionig to incraese communication efficiency," Artificial Life and Robotics, Vol.17, pp. 30-34, 2012.

[14] L. J. Villanueva-Rivera, B. C. Pijanowski, J. Doucette, and B. Pekin, "A primer of acoustic analysis for landscape ecologists," Landscape Ecology, Vol.26, pp. 1233-1246, 2011.

[15] Z. Chen and R. C. Maher, "Semi-automatic classification of bird vocaliza-tions using spectral peak tracks," The J. of Acoustical Society of America, Vol.120, pp. 2974-2984, 2006.
[16] O. R. Tachibana, N. Oosugi, and K. Okanoya, "Semi-automatic classification of birdsong elements using a linear support vector mahine," PLOS ONE, Vol.9, No.3, e92584, 2014.
[17] T. S. Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," Bird Conservation International, Vol.18, pp. 163-173, 2008.
[18] E. P. Kasten, P. K. McKinley, and S. H. Gage, "Ensamble extraction for extraction and detection of bird species," Ecological Informatics, Vol.5, No.3, pp. 153-166, 2010.
[19] L. Neal, F. Briggs, R. Raich, and X. Z. Fern, "Time-frequency segmentain of bird song in noisy acoustic environemnts," IEEE ICASSP, pp. 2012-2015, 2011.
[20] C. H. Lee, C. C. Han, and C. C. Chuang, "Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients," IEEE Trans. on audio, speech, and language processing, Vol.16, No.8, pp. 1541-1550, 2008.
[21] F. Briggs et al., "Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach," The J. of Acoustical Society of America., Vol.131, No.6, pp. 4640-4650, 2009.
[22] D. Stowell and M. D. Plumbley, "Automatic large-scale classificaiton of bird sounds is strongly improved by unsupervised feature learning," PeerJ, Vol.2, e488, 2014.

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:
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:
Fumiyuki Saito

Affiliation:
Senior Researcher, IDEA Consultants, Inc.

Address:
1-24-22 Nankokita, Suminoe-ku, Osaka-shi, Osaka 559-8519, Japan

Brief Biographical History:
2000 Received Master of Natural Science from Graduate School of Natural Science, Chiba University
2000- IDEA Consultants, Inc.

Membership in Academic Societies:
- Wild Bird Society of Japan
- Japan Society of Erosion Control Engineering

Name:
Tatsuyoshi Murate

Affiliation:
Chief of the Environmental Conservation Section, IDEA Consultants, Inc.

Address:
1-24-22 Nankokita, Suminoe-ku, Osaka-shi, Osaka 559-8519, Japan

Brief Biographical History:
1992- Graduated from Department of Biology, Faculty of Science, Okayama University
1992- IDEA Consultants, Inc.

Membership in Academic Societies:
- The Ornithological Society of Japan
- Asian Raptor Research and Conservation Network

Name:
Tomohisa Masuda

Affiliation:
Senior Researcher, IDEA Consultants, Inc.

Address:
1-5-12 Higashihama, Higashi-ku, Fukuoka City, Fukuoka 812-0055, Japan

Brief Biographical History:
1992 Received Master of Science from Graduate School of Science, Kyushu University
1998 Completed Ph.D. program without a degree, Faculty of Science, Kyushu University
2007- IDEA Consultants, Inc.

Main Works:
- S. Yamagishi, T. Masuda, and H. Rakotomanana, "A field guide to the birds of Madagascar," Kaiyusha Publishers Co., Ltd., 1997.

Membership in Academic Societies:
- Japan Bird Banding Association
- Wild Bird Society of Japan

Name:
Koichi Yamamoto

Affiliation:
Senior Researcher, IDEA Consultants, Inc.

Address:
1-24-22 Nankokita, Suminoe-ku, Osaka-shi, Osaka 559-8519, Japan

Brief Biographical History:
2001 Received Master of Natural Science from Graduate School of Natural Science, Chiba University
2001- IDEA Consultants, Inc.

Membership in Academic Societies:
- The Ornithological Society of Japan

Name:
Ryosuke Kojima

Affiliation:
Graduate School of Information Science and Engineering, Tokyo Institute of Technology

Address:
2-12-1-W8-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

Brief Biographical History:
2014 Received Master of Engineering in Computer Science from Graduate School of Information Science and Engineering, Tokyo Institute of Technology
2014- Doctoral program, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

Main Works:
- R. Kojima, O. Sugiyama, and K. Nakadai, "Multimodal Scene Understanding Framework and Its Application to Cooking Recognition," Applied Artificial Intelligence, Taylor & Francis, Vol.30, No.3, pp. 181-200, 2016.
- R. Kojima and T. Sato, "Goal and Plan Recognition via Parse Trees Using Prefix and Infix Probability Computation," Inductive Logic Programming, Springer, LNAI, Vol.9046, pp. 76-91, 2015.

Membership in Academic Societies:
- The Robotics Society of Japan (RSJ)
- The Japanese Society for Artificial Intelligence (JSAI)

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, K. Nakadai, H. and G. Okuno, "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, T. Yoshida, K. Nakamura, and K. Nakadai, "SLAM-based Online Calibration for Asynchronous Microphone Array," Advanced Robotics, Vol.26, No.17, pp. 1941-1965, 2012.
• R. Takeda, K. Nakadai, T. Takahashi, T. Ogata, and H. G. Okuno, "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, T. Takahashi, H. G. Okuno et al., "Design and Implementation of Robot Audition System "HARK"," Advanced Robotics, Vol.24, No.5-6, pp. 739-761, 2010.
• K. Nakadai, D. Matsuura, H. G. Okuno, and H. Tsujino, "Improvement of recognition of simultaneous speech signals using AV integration and scattering theory for humanoid robots," 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