https://doi.org/10.20965/jrm.2017.p0154

Paper:

Design and Assessment of Sound Source Localization System with a UAV-Embedded Microphone Array

Kotaro Hoshiba[*1], Osamu Sugiyama[*2], Akihide Nagamine[*3], Ryosuke Kojima[*4], Makoto Kumon[*5], and Kazuhiro Nakadai[*1,*6]

[*1]Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
E-mail: hoshiba@cyb.mei.titech.ac.jp
[*2]Kyoto University Hospital
54 Kawaharacho, Shogoin, Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
[*3]Department of Electrical and Electronic Engineering, School of Engineering, Tokyo Institute of Technology
[*4]Graduate School of Information Science and Engineering, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
[*5]Graduate School of Science and Technology, Kumamoto University
2-39-1 Kurokami, Chuo-ku, Kumamoto, Kumamoto 860-8555, Japan
[*6]Honda Research Institute Japan Co., Ltd.
8-1 Honcho, Wako, Saitama 351-0188, Japan
[Received July 24, 2016; accepted December 15, 2016]

We have studied on robot-audition-based sound source localization using a microphone array embedded on a UAV (unmanned aerial vehicle) to locate people who need assistance in a disaster-stricken area. A localization method with high robustness against noise and a small calculation cost have been proposed to solve a problem specific to the outdoor sound environment. In this paper, the proposed method is extended for practical use, a system based on the method is designed and implemented, and results of sound source localization conducted in the actual outdoor environment are shown. First, a 2.5-dimensional sound source localization method, which is a two-dimensional sound source localization plus distance estimation, is proposed. Then, the offline sound source localization system is structured using the proposed method, and the accuracy of the localization results is evaluated and discussed. As a result, the usability of the proposed extended method and newly developed three-dimensional visualization tool is confirmed, and a change in the detection accuracy for different types or distances of the sound source is found. Next, the sound source localization is conducted in real-time by extending the offline system to online to ensure that the detection performance of the offline system is kept in the online system. Moreover, the relationship between the parameters and detection accuracy is evaluated to localize only a target sound source. As a result, indices to determine an appropriate threshold are obtained and localization of a target sound source is realized at a designated accuracy.

Keywords: robot audition, sound source localization, multiple signal classification, actual environmental measurement, unmanned aerial vehicle

1. Introduction

Research on outdoor sound processing is important, as it can be applied to various fields, such as measurement. In the Impulsing Paradigm Change through Disruptive Technologies Program (ImPACT) of the Cabinet Office, the Tough Robotics Challenge was launched to develop remote autonomous robots, which can work robustly in an extreme disaster environment. The importance of the base technologies of outdoor robots is thus being recognized. The analysis of an outdoor sound environment is an important theme as extreme audition in the Tough Robotics Challenge.

We have been studying sound source localization using a microphone array embedded on a UAV (unmanned aerial vehicle) with an aim to locate people who need assistance in a disaster-stricken area, based on the robot audition technology we developed. The robot audition is a Japan-originated research field, where interaction with people is realized mostly using indoor robots having ears [1]. To hear with ear of robots, remote recognition of speech is needed. To achieve this, we need to handle various kinds of noises. Thus, we studied various functions such as sound source localization, sound source separation and speech recognition using microphone array processing [2–5]. Sound processing technology using the microphone array has been studied in different approaches [6–11]. Moreover, the developed robot audition technologies have been released to the public as open source software HARK (Honda Research Institute Japan Audition for Robots with Kyoto University).[1]

1. http://www.hark.jp/ [Accessed January 24, 2017]

1.1. Difference Between Indoor and Outdoor Sound Environment Analysis

The above-mentioned noise problem is qualitatively different for indoor and outdoor environments. Therefore, the approach to the problem is different even though the same noise suppression technology is used. In the indoor environment, it is necessary to consider not only the noise from the surroundings, but also the reverberation effect. Reverberation is a serious problem, particularly, in speech recognition, which is known to be less robust against reverberation. It is difficult to avoid reverberation in an ordinary room because the room comprises walls, a ceiling, a floor, and other objects that reflect sound. There are competitions in international conferences, such as the Reverb Challenge for reverberation suppression technology.[2] On the other hand, reverberation contains information from the sound environment of the room. For example, using reverberation, sound source localization can be performed, although the accuracy is not as high as the accuracy of the sound source distance estimation using azimuth and elevation angles [11]. In a general outdoor environment, except in special situations, we do not have to consider the effects of reverberation, which also indicates that sound source distance estimation is difficult in an outdoor environment. Moreover, noise from the surroundings changes dynamically over a large dynamic range. The spatial distribution of the sound speed is not uniform and changes over time due to changes of wind, humidity, and temperature. There are many noise sources which cannot be identified as a point sound source and that cannot be modeled.

1.2. Related Studies of Outdoor Sound Environment Analysis

We have studied sound source localization methods based on the MUSIC (multiple signal classification) method [12] to solve the above problem, specific to the outdoor sound environment. The sound source localization methods include a cross-correlation method, and DSBF (delay and sum beamforming) [8, 10]. The MUSIC method is advantageous in that its spatial resolution is higher than that of the other methods. However, it has problems about significant reduction of accuracy in a dynamically changing noise environment and large calculation cost in eigenvalue decomposition processing. Solutions to the problems have been studied. For example, we proposed the method that handled dynamically changing noise by extending the GEVD-MUSIC (MUSIC based on generalized eigenvalue decomposition) [6] method for indoor sound source localization, iGEVD-MUSIC (incremental GEVD-MUSIC) [4]. The GEVD-MUSIC method can accurately localize a sound source even in a noisy environment which cannot be modeled, because sound signals recorded in advance are used to estimate a noise correlation matrix that gives noise information. However, it is difficult to handle dynamically changing noise with this method. On the other hand, the estimation of the noise correlation matrix with the iGEVD-MUSIC method uses past sound signals before the target time period under the assumption that noise is stationary in a short-time. Thus, the noise correlation matrix can be dynamically estimated and the outdoor sound source localization performance can be significantly improved. To extend these methods and respond to a change of noises by a UAV, a method of using a Gaussian process to obtain status information of the UAV for dynamical estimation of a noise correlation matrix was proposed [13]. In addition, to reduce the calculation cost of GEVD, we proposed the method as an extension to the GSVD-MUSIC (MUSIC based on generalized singular value decomposition) method [7], iGSVD-MUSIC (incremental GSVD-MUSIC) [5]. This method can handle the estimation error of the noise correlation matrix using in combination with the CMS (correlation matrix scaling) method, which can scale the amplitude of the noise correlation matrix. With these methods, a speech source at a distance of approximately 15 m and a clear sound source, such as a whistle, at a distance of approximately 20 m could be accurately localized. In this way, the element technologies of the sound source localization technology have been developed in a robust form against the outdoor environment.

1.3. Problem and Approach

However, these technologies have the following practical problems.

1. Only one-dimensional sound source localization (azimuth angle) is performed, although the outdoor environment is three-dimensional.

2. There is no visualization tool to display sound source localization results in three dimensions, and it is difficult to see the results intuitively.

We attempt to solve these problems in this paper.

To solve the first problem, by extending the existing method to involve altitude localization, a robust method of sound source localization on an azimuth-elevation angle plane was proposed. Additionally, sound source localization providing information regarding the azimuth angle, elevation angle, and distance, was realized by estimating the distance to the sound source, based on the assumption that the source was located near the ground.

To solve the second problem, a new visualization tool, which realizes a three-dimensional display of a sound source localization result, was developed. Several visualization tools have been reported thus far showing the results of sound source localization performed by a remotely operated robot [14–16]. However, it was difficult for a third party to intuitively understand the sound source location, relative to the robot and surrounding environment, because visualization was made from a viewpoint of robot. In order to obtain a visualization tool with which the UAV situation and sound source localization results can be seen from a viewpoint of third party, a tool

2. http://reverb2014.dereverberation.com/ [Accessed January 24, 2017]

that displays three-dimensional position data from a sensor on the UAV and results of the sound source localization on Google Earth™ was developed. Because users can change the viewpoint freely on Google Earth™, they can intuitively check the surrounding environment, UAV, and sound source.

A system using these proposed methods and tools was structured, tested, and evaluated. First, an offline system was designed. The iGEVD-MUSIC method, which involves large calculation cost but has high performance, was used to structure the system. Sound signals obtained in the experiment were processed offline and used to evaluate the system. Next, the offline system was extended to an online system for real time processing. The iGSVD-MUSIC method, which requires less calculation cost, was used for the online system to perform the real time processing. Sound signals obtained in an actual environment were processed online, evaluated, and discussed.

2. Sound Source Localization Method

In this section, the proposed sound source localization method is described.

2.1. iGEVD-MUSIC and iGSVD-MUSIC Methods

As mentioned in Section 1.2, various MUSIC methods have been proposed. The iGEVD-MUSIC method, which achieves higher performance than the other MUSIC methods, is used in the offline system. The iGEVD-MUSIC method was developed by improving the GEVD-MUSIC method to realize successive noise correlation matrix estimation. With the improved method, one can perform robust sound source localization in a dynamically changing noise environment. The iGSVD-MUSIC method is used in the online system to ensure real time processing. The iGSVD-MUSIC method can localize a sound source with a small calculation cost, even in a dynamically changing noise environment.

The algorithm is described. M channel input sound signals of the f-th frame are Fourier transformed to $Z(\omega, f)$, from which a correlation matrix $R(\omega, f)$ is defined as follows:

$$R(\omega, f) = \frac{1}{T_R} \sum_{\tau=f}^{f+T_R-1} Z(\omega, \tau) Z^*(\omega, \tau). \quad \ldots \quad (1)$$

ω is the frequency bin number, T_R is the number of frames used for the correlation matrix calculation, and Z^* is a complex conjugate transpose of Z. Next, for f-th frame, the section of the length of T_N frames from the $f - f_s$-th frame is assumed to be a noise section, and the noise correlation matrix $K(\omega, f)$ is calculated.

$$K(\omega, f) = \frac{1}{T_N} \sum_{\tau=f-f_s-T_N}^{f+f_s} Z(\omega, \tau) Z^*(\omega, \tau) \quad \ldots \quad (2)$$

The GEVD-MUSIC method uses the noise correlation matrix calculated in advance from the given noise section, and hence, cannot respond to a dynamical change in noise. The iGEVD-MUSIC method and iGSVD-MUSIC method can estimate noise in each frame and is able to respond to a dynamical change in noise. The noise component can be whitened by multiplying K^{-1} to R from the left. The iGEVD-MUSIC method calculates eigenvectors through a GEVD of thus obtained $K^{-1}(\omega, f)R(\omega, f)$.

$$K^{-1}(\omega, f)R(\omega, f) = X(\omega, f)\Lambda(\omega, f)X^*(\omega, f) \quad (3)$$

$\Lambda(\omega, f)$ is a matrix with diagonal components that are eigenvalues in a descending order. $X(\omega, f)$ is a matrix containing eigenvectors corresponding to $\Lambda(\omega, f)$. Using X, and a transfer function, $G(\omega, \psi)$, corresponding to the sound source direction, ψ in the UAV coordinate system, the MUSIC space spectrum, $P(\omega, \psi, f)$, is calculated.

$$P(\omega, \psi, f) = \frac{|G^*(\omega, \psi)G(\omega, \psi)|}{\sum_{m=L+1}^{M} |G^*(\omega, \psi)x_m(\omega, \psi)|} \quad \ldots \quad (4)$$

L is the number of target sound sources, and x_m is the m-th eigenvector contained in X. The iGSVD-MUSIC method calculates singular vectors through the GSVD of $K^{-1}(\omega, f)R(\omega, f)$.

$$K^{-1}(\omega, f)R(\omega, f) = Y_l(\omega, f)\Sigma(\omega, f)Y_r^*(\omega, f) \quad (5)$$

$\Sigma(\omega, f)$ is a matrix with diagonal components that are singular values in a descending order. $Y_l(\omega, f)$ and $Y_r(\omega, f)$ are matrices containing singular vectors corresponding to $\Sigma(\omega, f)$. Then, the MUSIC space spectrum, $P(\omega, \psi, f)$, is calculated as in the iGEVD-MUSIC method.

$$P(\omega, \psi, f) = \frac{|G^*(\omega, \psi)G(\omega, \psi)|}{\sum_{m=L+1}^{M} |G^*(\omega, \psi)y_m(\omega, \psi)|} \quad \ldots \quad (6)$$

y_m is the m-th singular vector contained in Y_l. $P(\omega, \psi, f)$ thus obtained is average over ω direction to estimate the direction of the sound source.

$$\bar{P}(\psi, f) = \frac{1}{\omega_H - \omega_L + 1} \sum_{\omega=\omega_L}^{\omega_H} P(\omega, \psi, f) \quad \ldots \quad (7)$$

ω_H and ω_L are indices corresponding to the upper and lower limits of the used frequency bin, respectively. Threshold processing and peak detection are performed for $\bar{P}(\psi, f)$ and ψ of the obtained peak is detected as the sound source direction.

2.2. Two-Dimensional Sound Source Localization Including Elevation Angle

In general, the sound source direction ψ is presented only by the azimuth angle θ and this one-dimensional localization is sufficient for indoor use. However, for outdoor sound source localization by a UAV, localization of the elevation angle ϕ is also necessary. Therefore, in the proposed method, this one dimension is extended to two dimensions as follows:

$$\psi = (\theta, \phi). \quad \ldots \quad (8)$$

This definition does not lose the generality of the above MUSIC algorithm. In what follows, θ is defined with the front direction of the UAV being set to 0° and the back direction set to 180° (−180°). θ increases from −180° to 180°, counterclockwise, and the evaluation is made in this range. ϕ is defined with the vertical downward direction of the UAV being set to −90° and the horizontal direction set to 0°. The evaluation is made in the range of ϕ from −90° to 0°. The threshold processing and peak detection needs to be performed, not on the θ line, but on the θ-ϕ plane. In this method, the following single sound source was assumed and localized it by simply detecting a maximum value.

$$\Psi(f) = \text{argmax}_{\psi \in \{\psi | \bar{P}(\psi,f) \geq P_{th}\}} \bar{P}(\psi, f) \quad . \ . \ . \quad (9)$$

P_{th} is a threshold for the judgment of sound source.

2.3. Sound Source Distance Estimation

A method of estimating, not only the azimuth and elevation angles, but also distance to the sound source is described. The sound source direction is presented as $\Psi(f) = |\Theta(f), \Phi(f)|$ in the polar coordinate system, as in the previous section. This is because the direction is presented in a three-dimensional form in Cartesian coordinate system, consisting of the axes x, y, and z, and the sound source localization with azimuth and elevation angle information is often called three-dimensional sound source localization. However, because it contains only the azimuth and elevation angles, it is not actually three-dimensional sound source localization. For realization of true three-dimensional sound source localization, estimation of distance to the sound source is necessary. To analyze an outdoor sound environment, distance information would also need to be estimated to display the sound source on a map. However, because, as mentioned in the previous section, it is difficult to use reverberation, which is an important cue for estimation of distance, the sound source distance estimation is a difficult problem in an outdoor environment. This problem is circumvented by assuming that the sound source is located near the ground (at the height of a person's mouth). First, the sound source direction, which is obtained from the UAV coordinate system, is converted to the absolute coordinate system using the posture information of the UAV obtained from the navigation data, in order to derive a pair of azimuth and elevation angles $[A, E]$ in the absolute coordinate system. With the altitude of the UAV from the ground being expressed as h, and that of the sound source as h_{src}, the sound source distance can be given as follows:

$$D = \left| \frac{h - h_{src}}{\sin(E)} \right| . \quad (10)$$

Thus, the sound source position can be expressed as follows, if the center of the UAV is taken as origin:

$$P_s = [A, E, D] \quad \text{(in polar coordinate system)} \quad . \ (11)$$

$$= [D\cos(E)\cos(A), -D\cos(E)\sin(A), D\sin(E)] \quad \text{(in Cartesian coordinate system)} \quad . \ (12)$$

In this paper, a combination of the two-dimensional sound source localization and estimation of the distance information is called 2.5 dimensional sound source localization.

3. Offline Sound Source Localization in Actual Environment

First, an offline sound source localization system was structured using the proposed method, and then, evaluated and discussed experimental results of sound source localization performed in an actual environment.

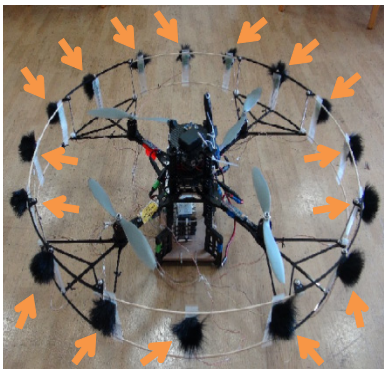
3.1. Measurement Situation

An outdoor evaluation experiment was performed. Twenty-one different sounds were generated from a speaker and localized these sound. A Pelican (AscTec) and Zion (enRoute) were used as the UAVs. The payload was 650 g and 4.0 kg, and flight time was 16 min and 30 min, respectively. The Pelican has a gyro, altitude sensor, GPS, acceleration sensor, and magnetic sensor to collect navigation data such as position, posture, velocity, and acceleration. As shown in **Figs. 1(a)** and **(b)**, a compact, light-weight microphone array consisting of 16 MEMS microphones is attached around or under the UAVs. As shown in **Fig. 1(c)**, a similar microphone array is attached around a helium-gas balloon, with which the same experiment is performed as the UAVs. The arrows in the figure indicate the position of the microphones. **Table 1** shows the measurement condition. The experiments were performed with two conditions, "Fixed" and "Flying," to record sound signals. In the "Fixed," the UAVs were fixed, even with the rotor spinning. In the "Flying," the UAVs flew and hovered. However, the balloon could not be completely fixed, as it was easily affected by wind. Compared to the "Fixed," the flying UAVs were largely impacted by the wind and had to respond to a dynamical change in the rotation sound of the rotor. In addition, the position and direction of the sound source could only be roughly estimated. **Fig. 2** shows the 21 different sound sources and volume levels used in the experiments. The sound was generated from wav files and volume levels are defined with the maximum value being 0 dB. The volume level is an indication, however, has no complete correlation with the ease of localization because the frequency characteristics of the sound sources are different from each other. The transfer function (G in Eq. (4)) used in the MUSIC method was derived, not from actual measurement, but from geometrical calculation.

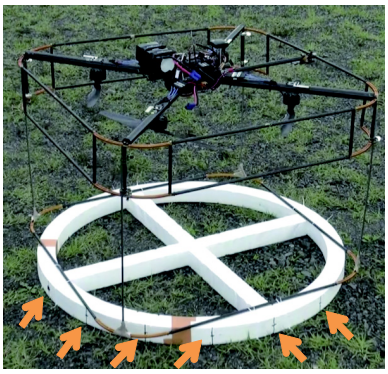
3.2. Structure of Offline System

The structured offline sound source localization system is described. The configuration is shown in **Fig. 3**. A multi-channel sound signal recorder, RASP-24[3] (System In Frontier), and microphone array were mounted on

3. http://www.sifi.co.jp/system/modules/pico2/index.php?content_id=4 [Accessed January 24, 2017]



(a) Pelican: A microphone array attached to a frame around the UAV.



(b) Zion: A microphone array on styrene foam attached beneath the UAV.

(c) Balloon: A microphone array attached around the balloon. Microphones are indicated by blue lamps.

Fig. 1. UAVs with microphone array.

Table 1. Experiment conditions. Altitude, distance, and angle are approximate values.

Label	UAV		Direction of sound source		Used sound source	
	Altitude [m]	Horizontal distance [m]	Azimuth angle [deg]	Elevation angle [deg]	Sound source type	Number of measurements (of each sound source)
Pelican fixed	0	3	0	0	21	10
Zion fixed	0	3	0	0–360 (with interval of 45)	1	10
Balloon fixed	0	3	0	65	20	10
Pelican flying A	5	3	60	0	7	3–10
Pelican flying B	5	5	45	0	7	3–10
Pelican flying C	5	10	27	0	7	3–10

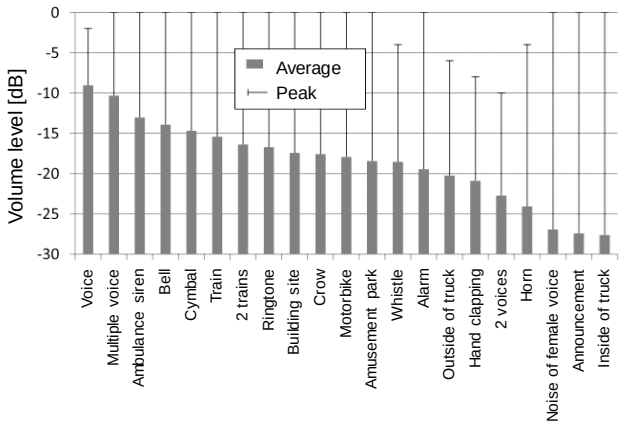


Fig. 2. Type and volume level of used sound sources.

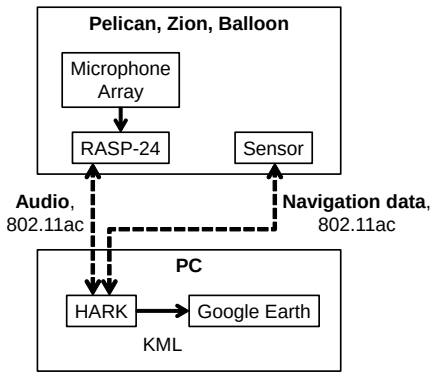


Fig. 3. Configuration of offline sound source localization system.

both the Pelican and Zion for the synchronous recording of 16 ch sound signals. For the balloon, RASP-ZX,[4] smaller and lighter than RASP-24, was employed for the synchronous recording of sound signals. The sound signals were recorded at a sampling frequency of 16 kHz, and quantization bit rate of 24 bits. The recorded sound signals and data from the sensors on the UAVs are transmitted through Wi-Fi (IEEE 802.11ac) to the processing PC. The received sound signals are processed in the PC for localization by the iGEVD-MUSIC method after the sound recording completes. HARK [17] was used for the algorithm implementation. One of the characteristics of this sound source localization system is the system can be applied to various microphone arrays by changing only the transfer function in Eq. (4), and this was achieved by HARK. With this versatile system, the sound source localization system can be applied for practical use. The sound source position is calculated in the absolute coordinate system by using the two-dimensional sound source

4. http://www.sifi.co.jp/system/modules/pico/index.php?content_id=36 [Accessed January 24, 2017]

localization data in the UAV's polar coordinate system and navigational data. The sound source localization results are converted into a form of KML (Keyhole Markup Language),[5] and the KML data of the localization results are then visualized on Google Earth™.

3.3. Evaluation Indices

The following three indices were used to evaluate the sound source localization.

- Index 1: Localization accuracy along different axes
- Index 2: Localization accuracy on UAV
- Index 3: Localization accuracy on sound source

Let N be the total number of sound sources, $C_{a_{th}}$ be the number of sound sources correctly localized within the azimuth or elevation angle, a_{th}, from the viewpoint of the UAV, S be the number of sound sources localized not located within the angle a_{th}, D be the number of sound sources not localized, and I be the number of the sources that were not actual sound sources and wrongly counted as sound sources. Index 1 is calculated by $(N-S-D-I)/N$. Because $C=N-S-D$, Index 1 becomes negative with large I. This index is similar to LAR (localization accuracy rate) [9] and its accuracy does not change depending on the distance between the UAV and sound source. In this experiment, a_{th} of the azimuth angle was set to 5° and that of the elevation angle was set to 10°.

Index 2 shows whether the localized point lies within an angle of b_{th} from the sound source direction, from the viewpoint of the UAV. Like Index 1, the accuracy of Index 2 does not change depending on the distance between the UAV and sound source. According to the LCR (localization correct rate) [9], the index is given by $C_{b_{th}}/N$ where $C_{b_{th}}$ is the number of sound sources correctly localized within the angle b_{th}. The localization is judged to be successful when the following two conditions are simultaneously satisfied, as indicated by the shaded area in **Fig. 4(a)**.

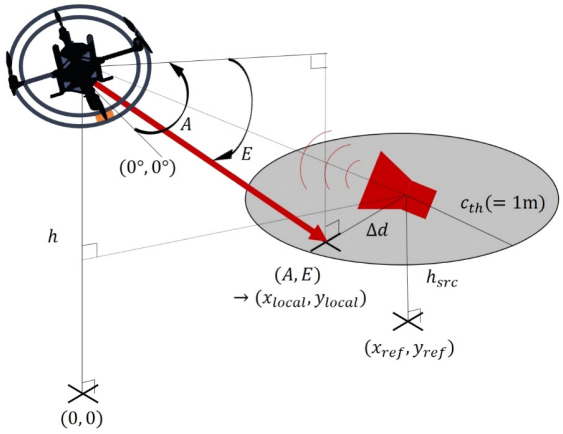
$$|A-A_{ref}| \leq b_{th} \quad \ldots\ldots\ldots\ldots\ldots (13)$$
$$|E-E_{ref}| \leq b_{th} \quad \ldots\ldots\ldots\ldots\ldots (14)$$

In this experiment, b_{th} was set to 10°.

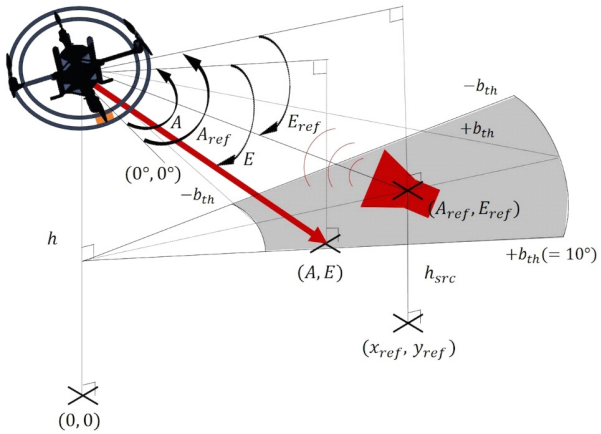
Index 3 shows whether the localized point lies within a distance of c_{th} from the sound source position. Like Index 2, Index 3 is given by $C_{c_{th}}/N$ where $C_{c_{th}}$ is the number of sound sources correctly localized within the distance c_{th}. The localization is judged to be successful when the following condition is satisfied, as indicated by the shaded area in **Fig. 4(b)**.

$$\Delta d = \sqrt{(x_{ref}-x_{local})^2+(y_{ref}-y_{local})^2} \leq c_{th} \quad (15)$$

In this experiment, c_{th} was set to 1 m. Because the sound source localization is created in the polar coordinate system, the accuracy of this index is low, when the distance



(a) Index 2



(b) Index 3

Fig. 4. Evaluation indices.

of the sound source from the UAV is long, even with the same sound localization. Moreover, because the reference value in **Table 1** was not very accurate, the value was not used as it was. Instead, the value was calibrated in the following manner. A histogram of the localization results was created and the central value of the histogram was used as a reference, if it was within ±20° from the value in **Table 1**.

3.4. Localization Results

The localization result of the offline system is evaluated. **Table 2** shows the calculation result of Index 1 with both the Pelican and balloon fixed in position. With the fixed Pelican, the sound localization was successful, even in the presence of the rotor sound. The ringtone localization performance was lower than the performance of the localization of the other sound sources. This could be because the ringtone has power concentrated at a specific frequency, and hence, tends to be buried in the rotor sound frequencies. On the other hand, the balloon showed lower localization performance than the Pelican, irrespective of no rotor sound. In particular, sound source localization in terms of the elevation angle could be obtained, however, the localized direction was largely different from the

5. https://developers.google.com/kml/ [Accessed January 24, 2017]

Table 2. Calculation results of Index 1 for balloon fixed and Pelican fixed [%].

No symbol: 500–2800 Hz. *: 2800–6000 Hz. –: Cannot be calculated.

		Voice	Multiple voice	Ambulance siren	Bell	Cymbal	Train	2 trains
Balloon	Azimuth	100	80	100	90	60	100	80
	Elevation	0	0	0	0	0	0	0
Pelican	Azimuth	100	100*	100	100*	100	100*	100
	Elevation	100	100*	100	100*	100	100*	100

		Ringtone	Building site	Crow	Motorbike	Amusement park	Whistle	Alarm
Balloon	Azimuth	60*	100	70	–	100	100	100
	Elevation	0*	0	0	–	0	0	0
Pelican	Azimuth	60*	100*	100	100*	100	100*	100
	Elevation	60*	100*	100	100*	100	100*	100

		Outside of truck	Hand clapping	2 voices	Horn	Noise of female voice	Announcement	Inside of truck
Balloon	Azimuth	70	60	80	50	70	80	70
	Elevation	0	0	0	0	0	0	0
Pelican	Azimuth	100	90*	100	100	100*	100	100*
	Elevation	100	90*	100	100	100*	100	100*

Table 3. Calculation results of Index 1 for Zion fixed [%].

	Azimuth	Elevation
−90°	90	90
−45°	100	100
−0°	100	100
45°	100	70
90°	100	90
180°	90	80

Table 4. Calculation results of Index 2 for Pelican in flight [%].

	Voice	Ambulance siren	Bell	Crow	Whistle	Horn	Announcement
Pelican flying A	–	86	100	100	90	60	0
Pelican flying B	70	–	–	–	100	100	–
Pelican flying C	80	–	–	–	90	60	–

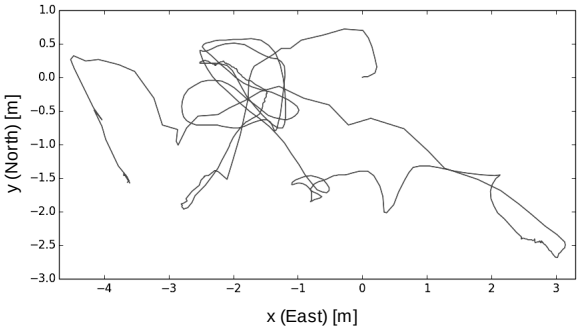
Table 5. Calculation results of Index 3 for Pelican in flight [%].

	Voice	Ambulance siren	Bell	Crow	Whistle	Horn	Announcement
Pelican flying A	–	86	100	100	70	80	0
Pelican flying B	40	–	–	–	90	50	–
Pelican flying C	20	–	–	–	0	20	–

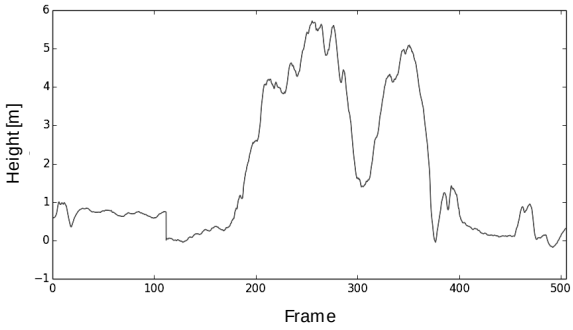
correct one. Some reasons for the poor localization performance included the inability to fix the balloon's position owing to the wind, as mentioned above, and the deformation of the microphone array layout attached to the surface of the balloon, which was deformed by the wind. The sound source localization performance of the balloon was expected to be better than that of the Pelican, because the balloon did not generate rotor sound. Therefore, the localization performance of the balloon should be enhanced by improving the microphone configuration and layout. **Table 3** shows the calculation result of Index 1 with the Zion fixed in position. The sound source localization performance of Zion was investigated for different sound source directions. The analysis of the elevation angle was limited in the range from $-45°$ to $0°$. The calculation result indicates that the localization could be performed, even in the presence of noise from the rotor. The performance changed a little depending on the direction, however, the navigation data analysis showed that the azimuth angle dependence of the localization performance was caused by wind. **Table 4** shows the calculation result of Index 2 with the Pelican in flight, and **Table 5** shows the calculation result of Index 3 under the same condition. The data show that localization is more difficult under the "Flying" condition than under the "Fixed" condition. The results for the whistle sound show that the Index 2 performance was high for any distance and Index 3 performance was lower as the distance became longer (A $\rightarrow$ C). The accuracy rate of Index 2 was high, even when the sound source distance was 10 m. However, Index 3 shows that the localization performance for the whistle sound was low at this distance. Namely, the 2.5 dimensional sound source localization including distance estimation is difficult at this distance. This indicates that step-by-step type active sound source localization is needed, estimating, first, only the two-dimensional sound source direction, approaching the sound source, and then, performing the 2.5 dimensional sound source localization. Because the distance to the sound source is short under condition A, there is no large difference between Index 2 and Index 3. Namely, localization of some sound source types was difficult, no matter which index was used. For example, the announcement sound could not be localized. This is because the SN (signal-to-noise) ratio of the announcement sound was small, owing to its small volume level, as shown in **Fig. 2**.

3.5. Visualization of Localization Results

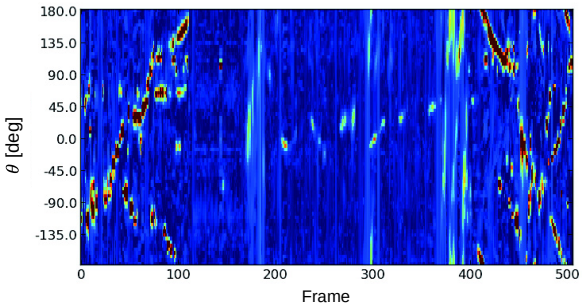
An experiment result under the "Flying" condition visualized with the visualization tool using Google Earth™ is shown. **Fig. 5** shows the experimentally obtained data. **Figs. 5(a)** and **(b)** show the UAV trace on the *x*-*y* plane and a temporal change in altitude, respectively. **Figs. 5(c)** and **(d)** show the MUSIC spectra of azimuth angle θ and elevation angle ϕ, respectively, of the recorded sound signals. The horizontal axis is the frame number, and the vertical axis is θ or ϕ. The power of the sound in each direction is shown on a color map. It is difficult to find three-dimensional motion of the UAV from **Figs. 5(a)** and **(b)**. From **Figs. 5(c)** and **(d)**, it could be seen that the



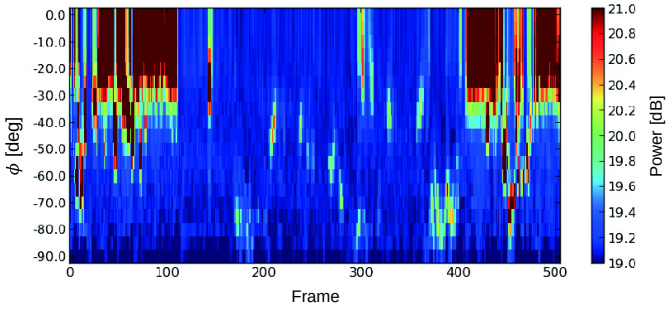
(a) Navigation data: x-y plane.



(b) Navigation data: Temporal change in altitude.



(c) MUSIC spectrum of azimuth angle θ.



(d) MUSIC spectrum of elevation angle ϕ.

Fig. 5. Obtained data. Three-dimensional motion and sound source direction of UAV cannot be intuitively recognized.

sound source localization was actually performed, however, it is difficult to intuitively find where and when the sound source was localized. **Fig. 6** shows snapshots of the data visualized on Google Earth^{TM}. The upper pictures are the photographs taken with a camera and lower ones are the visualized images on Google Earth^{TM}. The camera images did not show the location of the sound source, however, the images visualizing the results of the sound source localization show the three-dimensional motion of the UAV, and indicate the location of the sound source, a person, at the time of his/her speech. From the images, the usability of developed visualization tool can be confirmed.

4. Online Sound Source Localization in Actual Environment

The usability of the proposed offline system was confirmed in the previous section. Here, the offline system is extended to the online system and perform real-time sound source localization. An online sound source localization system was structured, and then, evaluated and discussed results of the sound source localization experiments conducted in an actual environment.

4.1. Measurement Condition

At the ImPACT Tough Robotics Challenge Test Meeting in June 2016, a demonstration of the online sound source localization was performed in an actual environment. Approximately 3 h of data obtained in the demonstration are evaluated. For the demonstration, the Bebop Drone (Parrot) was used as the UAV. The maximum payload was 200 g and flight time was 11 min. Compared to the Pelican or Zion, the Bebop Drone has a smaller payload, then, the body weight was reduced. The genuine battery weighed 115 g and had a capacity of 1200 mAh, however, the machine was modified to use a third party's lighter battery of a larger capacity (weight of 103 g and capacity of 1300 mAh). In the demonstration, however, we used an AC adaptor power supply for the Bebop Drone for prolonged operation. Navigation data could be obtained from the Bebop Drone, as from the Pelican, however, the data was not used as the measurement was performed by fixing the Bebop Drone. A microphone array of 8 MEMS microphones was mounted around the UAV as shown in **Fig. 7**. **Fig. 8** shows a demonstration layout of the sound source localization using the UAV. The demonstration was performed in a poster session area of the meeting. Because the exhibition area was next to the entrance, through which several people passed with significant noise. Moreover, another team performed a demonstration using an air compressor in an area next to our exhibition area, and the air compressor made significant noise. The Bebop Drone was fixed on a tripod at a height of approximately 1500 mm and the rotor spun at a constant 2000 rpm. A fixed sound source was placed at the lower front of the Bebop Drone and the sound of a barking dog was generated from a speaker. The audience spoke to the Bebop Drone from about 1–2 m distance in the area shown in **Fig. 8**. The sound source of the audi-

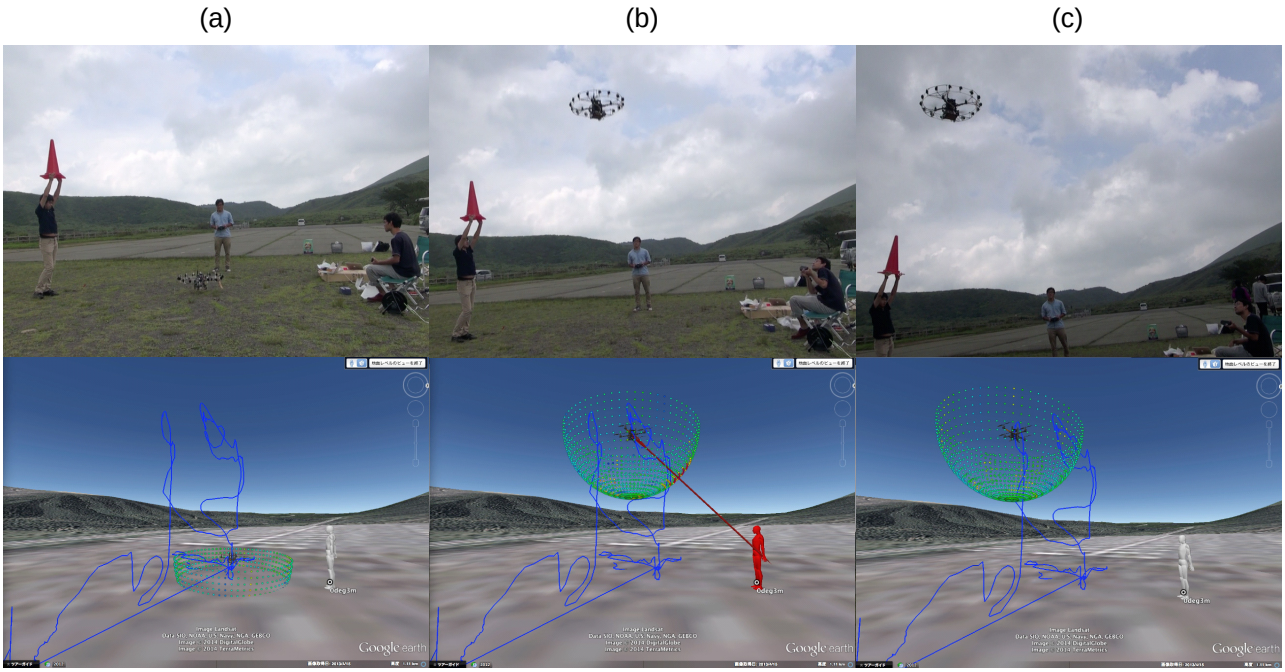


Fig. 6. Snapshots (upper: camera images, lower: Images created with visualization tool). (a) Starting flight, (b) speaking, (c) not speaking. Three-dimensional motion of the UAV and direction and position of a person, sound source, are displayed when the person speaks.

Fig. 7. Bebop Drone with microphone array. The arrows indicate the microphone array.

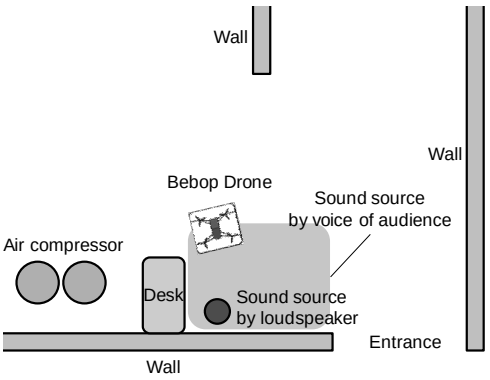


Fig. 8. Demonstration layout.

ence's voice and fixed sound source were used as target sound sources to localize.

4.2. Structure of Online System

The structured online sound source localization system is described. The configuration is shown in **Fig. 9**. Similar to the balloon experiment in the previous section, RASP-ZX and a microphone array were used for synchronous recording of 8 ch sound signals. The sound signals were recorded at a sampling frequency of 16 kHz and quantization bit rate of 24 bits. An AC adapter was used to supply power to the Bebop Drone for the long demonstration. To prevent the crossing of radio waves, RASP-ZX was wired through an Ethernet crossover cable to a processing PC, which processed the sound signals in real time using the iGSVD-MUSIC method. Similar to the offline system, HARK was used for the implementation of the algorithm. The sound source position in the absolute coordinate system was calculated from the obtained two-dimensional sound source localization data in the UAV polar coordinate system and navigation data, and transferred in a form of KML through the Ethernet crossover cable to the sound source localization result display tablet. On the tablet, the sound source localization results are displayed on Google Earth^{TM} using the KML data, similar to the offline system.

4.3. Results of Localization

The result of the sound source localization is shown. Only the azimuth angle θ and elevation angle ϕ of the Bebop Drone were used as evaluation parameters, as shown

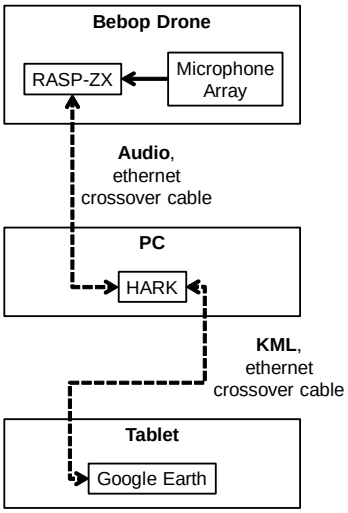


Fig. 9. Configuration of online sound source localization system.

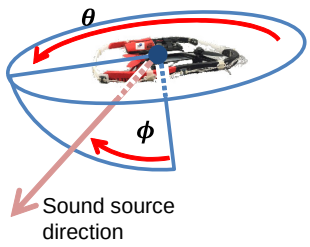


Fig. 10. Evaluation parameters.

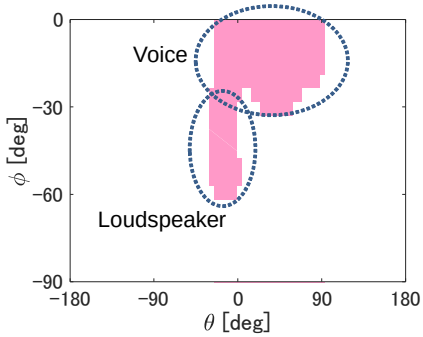


Fig. 11. Distribution of target sound source.

in **Fig. 10**, because the sound sources were located close to the drone. **Fig. 11** shows the sound source distribution area calculated from a rough positional relation between the Bebop Drone and sound sources. The area of the fixed sound source was at θ of approximately 0° and ϕ from approximately −60° to −30°, and the area of the audience's voice was at θ from approximately 0° to 90° and ϕ from approximately −30° to 0°. **Figs. 12(a)–(d)** show the MUSIC spectra obtained at different times. The horizontal axis is θ and vertical axis is ϕ. The power of the sound in each direction is shown on a color map. At the times of **Figs. 12(a)** and **(b)**, the proposed method suppressed stationary noise and detected strong power in the direction

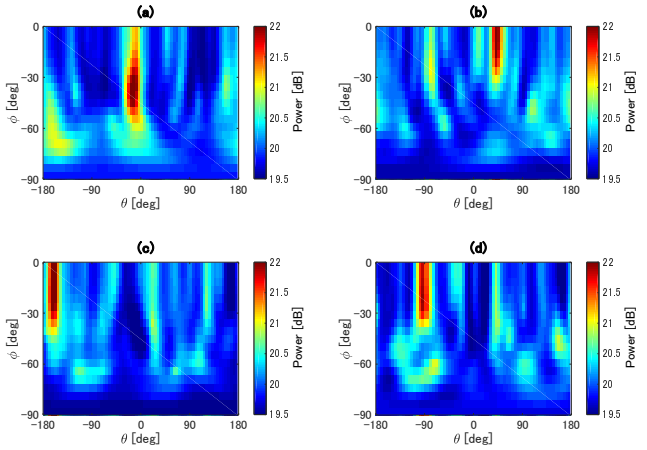


Fig. 12. MUSIC spectra at various times.

of the fixed sound source and audience's voice. At the times of **Figs. 12(c)** and **(d)**, the method could also detect the direction of the noise behind the Bebop Drone and air compressor.

4.4. Change in Detection Accuracy by Parameters

The localization result of the online system is evaluated. As in the previous section, various sound sources, other than the target sound source, could be detected in the measurement. However, these sources were not target sources. The method of identifying the sound source type using deep learning was proposed [18], and this method could be used to extract only the target sound source. However, implementation of the online sound source localization will be a future problem.

In this paper, sounds, other than the target sound, are excluded by adjusting the parameters, including the threshold and detection area. In the offline system, the threshold or detection area can be determined while checking all obtained data. However, in the online system, those parameters have to be determined for unknown data. In order to have a guideline of appropriate parameter design, a change in the detection accuracy by changing the parameters was evaluated. Because the location of the target sound source is known in this measurement condition, the target sound source direction was estimated by setting the threshold within a specific angle range from the MUSIC spectrum, shown in the previous section. **Fig. 13** shows the histograms of the sound source estimation in each direction with different thresholds. **Figs. 13(a)**, **(b)**, **(c)** and **(d)** show the results with the threshold being set to 21.0 dB, 21.3 dB, 21.6 dB, and 22.0 dB. As shown in **Fig. 12**, the horizontal axis is θ and vertical axis is ϕ. In the figures, the frequencies of localizing sound sources in each directions is presented in a log scale on a color map. When the threshold is low, as shown in **Fig. 13(a)**, the sound sources in the directions other than the target sound source direction, were often localized. With a higher threshold, the sound sources in the directions other than the target sound source direction, were localized less fre-

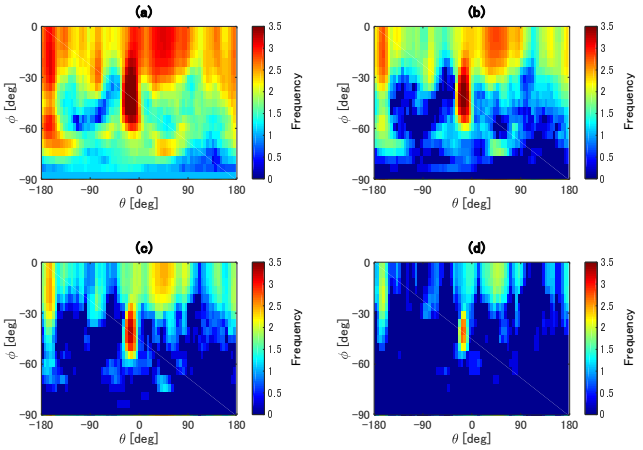


Fig. 13. Histograms of sound source estimation in each direction with threshold: (a) 21.0 dB, (b) 21.3 dB, (c) 21.6 dB, and (d) 22.0 dB.

quently. It was found that not only limiting the detection area, but also setting an appropriate threshold, was effective in distinguishing target sound and non-target sound. Thus, we discuss the relation between the threshold and detection accuracy. The miss detection probability and false detection probability are evaluated to simplify the discussion. From the definition in Section 3.3, the miss detection probability is given by $(N-C)/N$, and false detection probability by $(S+I)/(C+S+I)$. In this measurement, exact direction of the sound source cannot be known because the sound source direction changed over time, and no calibration was performed. Thus, the miss detection probability and false detection probability was calculated using the sound source detected in the area shown in **Fig. 8** as the correct one. **Fig. 14** shows the miss detection probability of the target sound source. The miss detection probability was calculated by successively changing the threshold at an interval of 0.01. The horizontal axis is the threshold and vertical axis is the miss detection probability. When the threshold was 21.0 dB, the miss detection probability was almost 0%. When the threshold was 21.5 dB, the miss detection probability was approximately 80%. The false detection probability of the target sound source is shown in **Fig. 15**. Similar to **Fig. 14**, the false detection probability was calculated by successively changing the threshold at an interval of 0.01. The horizontal axis is the threshold and vertical axis is the false detection probability. The dotted line presents the false detection probability calculated from the result of the entire area detection, and the solid line is the false detection probability calculated from the result of the front area (θ: -90° to 90°). The false detection probability decreased by approximately 30% when the detection area was limited to the front side rather than when it is not. It is therefore effective to limit the detection area. Further, it can be seen that the false detection probability decreased when the threshold increased. Thus, the miss detection probability and false detection probability is a tradeoff relation. In particular, the miss detection probability changes significantly with the threshold, and hence, the threshold should be set in the order of 10^{-2} dB.

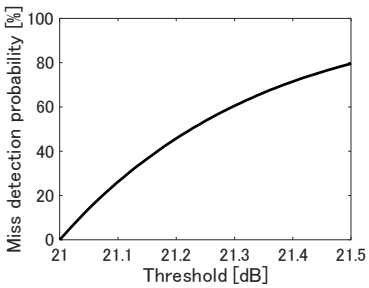


Fig. 14. Miss detection probability of target sound source.

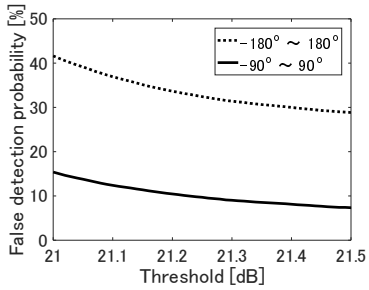


Fig. 15. False detection probability of target sound source.

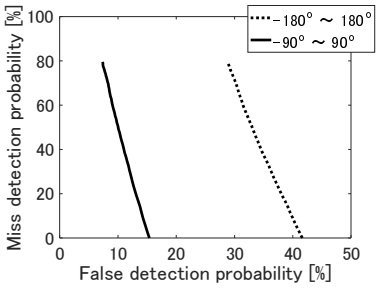


Fig. 16. DET curve.

4.5. Discussions

The parameter setting is discussed. The DET (detection error tradeoff) curve [19] was created from **Figs. 14** and **15**. The DET curve presents a change in the miss detection probability and false detection probability by changing the threshold, with the horizontal axis being the false detection probability and vertical axis being the miss detection probability. The curve is important to set an appropriate threshold. The DET curve of this measurement environment is shown in **Fig. 16**. The dotted line is a DET curve created from the result of the entire area detection and solid line is a DET curve created from the result of the front area detection. A threshold was set using these DET curves as indices. Because the false detection probability and miss detection probability are in a tradeoff relation, the threshold should be set difference which probability attach weight. Reduction of the miss detection probability is important because the aim of this study is to locate a

victim in a disaster-stricken place. In this demonstration, the detection area was limited to the front area, and the threshold was set such that the miss detection probability was approximately 5% and false detection probability was approximately 15%.

However, the following problems could arise when the system is used in a disaster-stricken area.

- The threshold and other parameters have to be appropriately set in real time as the sound recording condition dynamically changes in actual flying situations.
- In the demonstration, the target sound source and non-target sound source were located in different directions, and the directions were known. However, in actual situations, sound from a target sound source and sound from a non-target sound source could come from any direction.
- When a victim's voice is detected from above, the distance to the victim would be much larger than the distance used in the demonstration, and the SN ratio of the same level as in the demonstration would not be ensured owing to the presence of debris and the victim's health condition.

Solutions to these problems, as well as development of an online speech recognition system must be studied in future work.

5. Conclusions

In this paper, proposed method was extended to obtain a more practical one, a system was designed and implemented, and sound source localization experiments were then conducted in an actual environment. First, a 2.5 dimensional sound source localization method was proposed and used the method to structure an offline sound source localization system. The accuracy of the localization results was then evaluated and discussed. As a result, the usability of the extended proposed method and newly developed three-dimensional visualization tool could be confirmed. Further, a change in the detection accuracy owing to the type and distance of sound source could be found. Next, the offline system was extended to an online system and performed real-time sound source localization. The sufficient localization performance level could be confirmed, as in the offline case. Then, localization of only a target sound source was studied. In the real-time sound source localization, it is necessary to set a threshold and other parameters to unknown data. Therefore, the relation between the parameters and detection accuracy was evaluated. As a result, it was found that an appropriate threshold could be determined from a DET curve calculated from the miss detection probability and false detection probability, and the target sound source could be detected at a desired target accuracy. However, problems with the sound source localization in actual environments were found, and these problems should be examined in future work.

Acknowledgements

The authors would like to thank Keita Okutani, Takuma Ohata and Satoshi Uemura at Tokyo Institute of Technology, and Keisuke Nakamura at Honda Research Institute Japan for their support. This work was supported by KAKENHI No.24220006, 16H02884, 16K00294, and ImPACT Tough Robotics Challenge.

References:

[1] K. Nakadai, T. Lourens, H. G. Okuno, and H. Kitano, "Active audition for humanoid," Proc. of 17th National Conf. on Artificial Intelligence (AAAI-2000), pp. 832-839, 2000.
[2] S. Yamamoto, K. Nakadai, M. Nakano, H. Tsujino, J. M. Valin, K. Komatani, T. Ogata, and H. G. Okuno, "Design and implementation of a robot audition system for automatic speech recognition of simultaneous speech," Proc. of the 2007 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU-2007), pp. 111-116, 2007.
[3] H. Nakajima, K. Nakadai, Y. Hasegawa, and H. Tsujino, "Blind source separation with parameter-free adaptive step-size method for robot audition," IEEE Trans. on Audio, Speech, and Language Processing, Vol.18, No.6, pp. 1476-1485, 2010.
[4] K. Okutani, T. Yoshida, K. Nakamura, and K. Nakadai, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," Proc. of the IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), pp. 3288-3293, 2012.
[5] T. Ohata, K. Nakamura, T. Mizumoto, T. Tezuka, and K. Nakadai, "Improvement in outdoor sound source detection using a quadrotor-embedded microphone array," Proc. of the IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), pp. 1902-1907, 2014.
[6] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsujino, "Intelligent sound source localization for dynamic environments," Proc. of the IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), pp. 664-669, 2009.
[7] K. Nakamura, K. Nakadai, and G. Ince, "Real-time super-resolution Sound Source Localization for robots," Proc. of the IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), pp. 694-699, 2012.
[8] M. Basiri, F. Schill, P. U. Lima, and D. Floreano, "Robust acoustic source localization of emergency signals from Micro Air Vehicles," Proc. of the IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), pp. 4737-4742, 2012.
[9] Y. Bando, T. Mizumoto, K. Itoyama, K. Nakadai, and H. G. Okuno, "Posture estimation of hose-shaped robot using microphone array localization," Proc. of the IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), pp. 3446-3451, 2013.
[10] Y. Sasaki, N. Hatao, K. Yoshii, and S. Kagami, "Nested iGMM recognition and multiple hypothesis tracking of moving sound sources for mobile robot audition," Proc. of the IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), pp. 3930-3936, 2013.
[11] K. Niwa, S. Esaki, Y. Hioka, T. Nishino, and K. Takeda, "An Estimation Method of Distance between Each Sound Source and Microphone Array Utilizing Eigenvalue Distribution of Spatial Correlation Matrix," IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, Vol.J97-A, No.2, pp. 68-76, 2014 (in Japanese).
[12] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. on Antennas and Propagation, Vol.34, No.3, pp. 276-280, 1986.
[13] K. Furukawa, K. Okutani, K. Nagira, T. Otsuka, K. Itoyama, K. Nakadai, and H. G. Okuno, "Noise correlation matrix estimation for improving sound source localization by multirotor UAV," Proc. of the IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), pp. 3943-3948, 2013.
[14] Y. Sasaki, S. Masunaga, S. Thompson, S. Kagami, and H. Mizoguchi, "Sound Localization and Separation for Mobile Robot Tele-Operation by Tri-Concentric Microphone Array," J. of Robotics and Mechatronics, Vol.19, No.3, pp. 281-289, 2007.
[15] Y. Kubota, M. Yoshida, K. Komatani, T. Ogata, and H. G. Okuno, "Design and Implementation of 3D Auditory Scene Visualizer towards Auditory Awareness with Face Tracking," Proc. of the Tenth IEEE Int. Symposium on Multimedia (ISM), pp. 468-476, 2008.
[16] T. Mizumoto, K. Nakadai, T, Yoshida, R. Takeda, T. Otsuka, T. Takahashi, and H. G. Okuno, "Design and Implementation of Selectable Sound Separation on the Texai Telepresence System using HARK," Proc. of the IEEE Int. Conf. on Robots and Automation (ICRA), pp. 2130-2137, 2011.

[17] 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, No.5-6, pp. 739-761, 2010.
[18] S. Uemura, O. Sugiyama, R. Kojima, and K. Nakadai, "Outdoor Acoustic Event Identification using Sound Source Separation and Deep Learning with a Quadrotor-Embedded Microphone Array," Proc. of the 6th Int. Conf. on Advanced Mechatronics, pp. 329-330, 2015.
[19] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," Proc. of the Fifth European Conf. on Speech Communication and Technology, pp. 1895-1898, 1997.

Name:
Kotaro Hoshiba

Affiliation:
Researcher, Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology

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

Brief Biographical History:
2016 Received Ph.D. (Engineering), Department of Mechanical and Control Engineering, Graduate School of Science and Engineering, Tokyo Institute of Technology (Major: Acoustic signal processing, Acoustic imaging)
2016- Researcher, Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology

Main Works:
• K. Hoshiba, S. Hirata, and H. Hachiya, "High accuracy measurement of small movement on object using airborne ultrasound," Japanese J. of Applied Physics, Vol.52, No.7, 07HC15-1-6, July 2013.
• K. Hoshiba, S. Hirata, and H. Hachiya, "Measurement of ultrasound transmission attenuation characteristics of canvas fabric," Acoustical Science and Technology, Vol.36, No.2, pp. 171-174, Feb. 2015.

Membership in Academic Societies:
• The Acoustical Society of Japan (ASJ)
• The Robotics Society of Japan (RSJ)
• Information Processing Society of Japan (IPSJ)

Name:
Osamu Sugiyama

Affiliation:
Kyoto University Hospital

Address:
54 Kawaharacho, Shogoin, Sakyo-ku, Kyoto City 606-8507, Japan

Brief Biographical History:
2007-2009 SONY
2009-2013 Advanced Telecommunications Research Institute International
2013-2016 Kyoto University and Tokyo Institute of Technology
2016- Kyoto University Hospital

Main Works:
• O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, "Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model," Connection Science (Special issues on android science), Vol.18, No.4, pp. 379-402, 2006.

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

Name:
Akihide Nagamine

Affiliation:
Department of Electrical and Electronic Engineering, School of Engineering, Tokyo Institute of Technology

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

Brief Biographical History:
2016- Department of Electrical and Electronic Engineering, School of Engineering, Tokyo Institute of Technology

Membership in Academic Societies:
• The Japan Society of Applied Physics (JSAP)

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, 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:
Makoto Kumon

Affiliation:
Kumamoto University

Address:
2-39-1 Kurokami, Chuo-ku, Kumamoto, Kumamoto 860-8555, Japan

Brief Biographical History:
2000-2005 Assistant Professor, Kumamoto University
2006- Associate Professor, Kumamoto University

Main Works:
• M. Kumon and S. Uozumi, "Binaural Localization for a Mobile Sound Source," J. of Biomechanical Science and Engineering, Vol.6, No.1, pp. 26-39, 2011.

Membership in Academic Societies:
• The Robotics Society of Japan (RSJ)
• The Institute of Electrical and Electronics Engineers (IEEE)
• The Japan Society of Mechanical Engineers (JSME)

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)