2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Posture Estimation of Hose-Shaped Robot using Microphone Array Localization

Yoshiaki Bando, Takeshi Mizumoto, Katsutoshi Itoyama, Kazuhiro Nakadai, and Hiroshi G. Okuno

Abstract— **This paper presents a posture estimation of hose-shaped robot using microphone array localization. The hose-shaped robots, one of major rescue robots, have problems with navigation because their posture is too flexible for a remote operator to control to go as far as desired. For navigational and mission usability, the posture estimation of the hose-shaped robot is essential. We developed a posture estimation method with a microphone array and small loudspeakers equipped on the hose-shaped robot. Our method consists of two steps: (1) playing a known sound from the loudspeaker one-by-one, and (2) estimating the microphone positions on the hose-shaped robot instead of estimating the posture directly. We designed a time difference of arrival (TDOA) estimation method to be robust against directional noise and implemented a prototype system using a posture model of the hose-shaped robot and an Extended Kalman Filter (EKF). The validity of our approach is evaluated by the experiments with both signals recorded in an anechoic chamber and simulated data.**

I. Introduction

Recent disasters, either natural, artificial, or combined, have demanded robots for search-and-rescue missions because victims in the aftermath of such disasters are widely distributed and such sites are too dangerous or difficult to deploy human rescue teams. Based on the experience of nine ground robots at five incidents in the United States, Robin Murphy, one of the most active rescue robot researchers, pointed out that two types of usability should be improved to help with disasters; *navigational* and *mission* [1]. Navigational usability focuses on navigation of rescue robots. This is often hampered by ineffective locomotion and lack of sensing for control. Mission usability focuses on obtaining sensor data and providing information to remote operators. This is often inhibited by mediation effects, flawed sensor systems and poorly designed displays. She proposed potential solutions including additional proprioceptive and exproprioceptive sensors, increased autonomy, and multi-modal displays.

One modality that has not been well investigated is *sound*. In search-and-rescue missions by human rescue teams, sound is one of the crucial cues in searching for victims. Robot audition is used to enhance the user interface to help a remote operator to recognize and understand the auditory situation with sound source localization and separation [2]. A telepresence robot called *Texai* of Willow Garage was equipped with robot audition software *HARK* [3] and displayed sound source directions superposed on a remote display. This system demonstrated the feasibility of an auditory display for the telepresence robot.

Y. Bando, T. Mizumoto, K. Itoyama and H. G. Okuno are with the Graduate School of Informatics, Kyoto University, Sakyo, Kyoto, 606–8501, Japan. yoshiaki@zeus.kuis.kyoto-u.ac.jp, {mizumoto, itoyama, okuno}@kuis.kyoto-u.ac.jp
K. Nakadai is with Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako, Saitama 351–0114, Japan. nakadai@jp.honda-ri.com

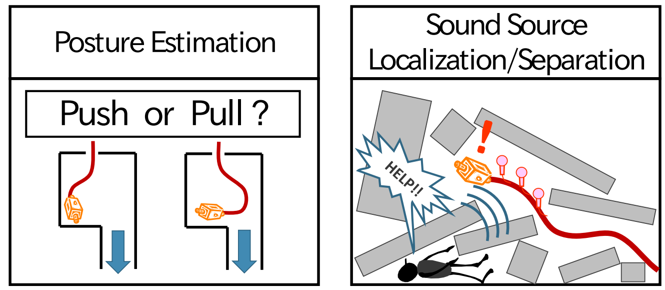


Fig. 1. Posture estimation, and sound source localization and separation are essential for improving navigational and mission usability.

Hose-shaped robots with a video camera on the tip have problems with navigational usability because their posture is too flexible for a remote operator to control to go as far as desired. Kitagawa et al. [4] developed Active-Hose, a hose-shaped robot that could move forward and change directions with small wheels and a timing belt. This robot can drive into narrow gaps in collapsed buildings which conventional large snake-like robots [5][6] cannot drive into. Hatazaki et al. [7] developed an active scope camera (ASC), a hose-shaped robot with vibrators and cilia. With its vibrating cilia set on whole surface of the robot, the ASC could drive into narrow gaps as long as the surface of the robot touches on the ground. In fact, the ASC was used for a search-and-rescue mission in Jacksonville, Florida, USA in 2008 [8]. Namari et al. [9] developed a tube-shaped ASC to enhance the mobility and implement some functions for practical use.

The posture estimation of the hose-shaped robot is essential for navigational and mission usability because of two reasons: First, the unexpected bending on the hose may make it difficult for an operator to control the robot as desired. Second, tip localization enables remote operators to estimate victim locations through remote images.

Ishikura et al. [10] reported the tip localization method using the temporal difference of tip camera visions. They estimated the movement of the tip camera by extracting similar points between images. This method often failed to extract similar points due to excessive brightness canceled by the tip light or the image having few characteristics, as with a smooth wall. Ishikura et al. [11] then estimated a ASC posture by using inertial sensors and magnetic sensors. They adopted a flexible dynamics model of the ASC and estimated its posture using an Unscented Kalman Filter (UKF). The estimation performance was often deteriorated because of its vibrators.

Robot audition may improve navigational usability. A hose-shaped robot with a set of microphones, or a microphone array, would be expected to improve both navigational and mission usability if it can localize microphone positions by sound generated by itself. Localizing microphone positions will improve the performance of posture estimation of a hose-shaped robot, which leads to enhance navigational usability. In addition, a microphone array with its configuration of microphones can localize sound sources, which leads to enhance mission usability.

This paper proposes a posture estimation of a hose-shaped robot using sound. The key idea is that we estimate the microphone locations on the hose-shaped robot instead of estimating the posture directly. Microphone array localization is estimated by the time difference of arrival (TDOA) between a pair of microphones [12][13]. As a preliminary experiment, instead of using the sound of electric motors, a set of small loudspeakers are used in this paper to verify whether sound is promising. The posture estimation consists of two steps: (1) playing a known sound from the loudspeaker one-by-one and (2) estimating the posture using the TDOA.

The rest of this paper is organized as follows: Section II reviews related works of microphone array localization. Section III introduces a mockup of the hose-shaped robot and defines the problem statement. Section IV describes the posture estimation method. Section V presents the results of an evaluation of our method. Section VI summarizes the paper and states our future work.

II. Related Works of Microphone Array Localization

Microphone array localization has been actively studied. We review these studies here for building a posture estimation using the TDOA.

Ono et al. [12] defined "Blind Alignment Problem" to estimate sound source locations, microphone locations, and microphone clock differences by using recorded signals of handclap sounds. They reveal the necessary condition of the number of microphones and handclaps for this problem, and showed a solution. They evaluated its performance by simulating the recorded signal of handclap sounds. Their method solves Blind Alignment Problem through an auxiliary function approach with the TDOA of handclap sound. Since their method was designed for offline operation, the calculation cost was extremely extensive, making it difficult to calculate in real time.

Miura et al. [13] reported an online microphone array localization method based on Simultaneous Localization and Mapping (SLAM) framework using a man who walks around a robot while clapping. They also used the TDOA of handclap sound. SLAM is a framework to estimate the self-location and landmark locations simultaneously. They applied the SLAM framework considering the man location as self-location and the microphone locations and clock differences as landmark. This method estimates the moving self-location and the landmarks with correcting the error from the TDOA of handclap sound using Extended Kalman Filter (EKF). This method works online because the man assumption lowers calculation costs. However, it is not practical for use with the hose-shaped robots due to the impossibility of a moving sound source in the collapsed buildings.

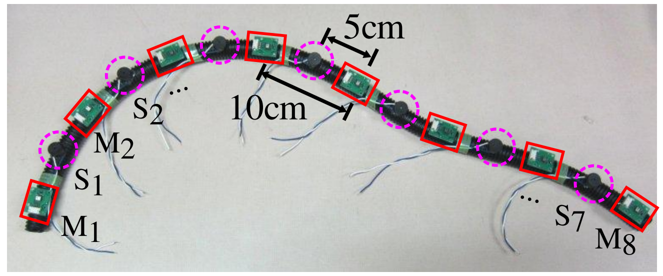


Fig. 2. Mockup of the hose-shaped robot. Squares and circles indicate microphones and loudspeakers, respectively.

Our method uses a limitation of the hose-shaped robot posture. The shape of the hose restricts the location of the microphones and loudspeakers. We import this limitation into our state space model and estimate the posture using EKF.

III. Problem Statement of Posture Estimation

In this section, we present a mockup of the hose-shaped robot, discuss the challenges with the TDOA estimation, and define the problem statement.

A. Description of the mockup hose-shaped robot

Figure 2 shows the mockup of the hose-shaped robot. The hose was corrugate tube made from polypropylene (inside diameter of 15 mm). We fixed microphones and small loudspeakers to the hose using double-sided tape. Microphones were placed at intervals of 10 cm, and loudspeakers were placed between neighboring microphones. We use a small (20 mm $\times$ 25 mm) microphones manufactured by micro-electro mechanical system (MEMS), and small loudspeakers with a radius of 14 mm made from a piezoelectric device.

Let N be the number of microphones ($N = 8$). Since loudspeakers are placed between microphones, the number of loudspeakers is $N - 1$. Let $\boldsymbol{M}_m, (m = 1 \cdots N)$ be the coordinate of the m-th microphone and $\boldsymbol{S}_l, (l = 1 \cdots N-1)$ be the coordinate of l-th loudspeaker. The other notations are summarized in Table I.

B. Challenges of TDOA estimation

In the TDOA estimation for the hose-shaped robot, the sound environment has four challenges:

1) **External noise**: The hose-shaped robot is used outside which includes numerous directional noises.
2) **Reverberation**: The hose-shaped robot is inserted in a narrow space, therefore, the reverberation time is long which makes sound processing difficult.
3) **Obstacles**: In a narrow space, recording a played sound directly is difficult when the space is canned.

TABLE I
NOTATIONS

t	Time
k	Measurement index
N	Number of microphones
m, n	Microphone index: $m, n \in [1, \cdots, N], m \neq n$
l	Loudspeaker index: $l \in [1, \cdots, N-1]$
$\boldsymbol{M}_m, \boldsymbol{M}_n \in \mathbb{R}^2$	Coordinate of m-th and n-th microphone
$\boldsymbol{S}_l \in \mathbb{R}^2$	Coordinate of l-th loudspeaker
$m_n(t), m_m(t)$	Signal of n-th and m-th microphone
$s(t)$	Known test signal
$\tau^l_{m \to n}$	TDOA between m-th and n-th signal from l-sth loudspeaker
$\boldsymbol{x}_k \in \mathbb{R}^{2N-3}$	State vector of the hose-shaped robot posture
c	Speed of sound
L	Distance between microphones

4) **Covered microphone**: The hose often rotates since its cross-section is circle. Therefore, some microphones can be under the hose. Such microphones are unreliable.

As a first step, we tackle the first challenge because the noise problem is inevitable. We assume that the hose-shaped robot is stationary during estimation, that all microphones are turned up, and that the hose curves on a 2D surface.

C. Problem statement

Problem statement

Input:
Synchronized N channel audio recording of a known sound
Output:
Locations of microphones
Assumptions:
(1) The microphones and loudspeakers are on the hose-shaped robot.
(2) Played signal, $s(t)$, and the index of the playing loudspeaker, l, are known.
(3) $\boldsymbol{M}_1$ and $\boldsymbol{S}_1$ are known.

The input is used as a clue for the TDOA estimation. The output is the microphone locations that represents the posture of the hose-shaped robot because we assume that the microphones are on the hose in the first assumption. The second assumption holds because we can control the timing of the sound. The third assumption avoids the ambiguity of the rotation and translation of posture estimation because the TDOA contains only the relative microphone position.

IV. EKF-BASED ONLINE POSTURE ESTIMATION

In this section, we describe the over view of our method, present the TDOA estimation, formulate the state space model of the robot posture, and describe the EKF for our estimation.

A. Overview

Our method estimates the robot posture by measuring the TDOA from the loudspeaker and solves the problem using an EKF. Alg.1 shows the procedure of one iteration.

Algorithm 1 The procedure of one iteration

```
for l = 1 → N − 1 do
    Start recording with N ch microphones
    Play the test signal from the l-th loudspeaker
    Stop recording

    Estimate the TDOA of the test signal
    Correct current state to minimize the error
end for
```

For this algorithm, we build a state space model of the hose-shaped robot posture, and TDOA estimation method.

B. TDOA Estimation

In this section, we describe a TDOA estimation using a sweep signal as the test signal and a cross correlation.

1) Design of test signal: The conventional microphone array localization methods [12][13] used hand clap sound which is a kind of impulsive sound. Since impulsive sound is a short in duration, and, the cross correlation has a sharp peak, meaning that, impulsive sound can make the TDOA estimation to be robust against the noises.

Since it is difficult to play an impulsive sounds with a small loudspeaker due to its limited frequency characteristics, we use a sweep signal as the test signal. The sweep signal is a sound whose frequency increases linearly. The frequency of a sweep signal is changing all the times, the resemblance sequence in sweep signal is few, so its cross correlation has a sharp peak. As a result, it is robust against directional noise. The sweep signal that sweeps from f_s Hz to f_e Hz during T second is represented as follows:

$$s(t) = \sin\left(2\pi \int_0^t \left(\frac{f_e - f_s}{T} t + f_s\right)\right) \tag{1}$$

$$= \sin\left(\pi \left(\frac{f_e - f_s}{T} t^2 + 2 f_s t\right)\right) \tag{2}$$

2) Cross correlation: Ono et al. [12] developed a TDOA estimation method by finding the maximum cross correlation between recorded signals. When a loud directional noise disturbs the recording, the method outputs the TDOA of the noise because the cross correlation is dominated by the loud noise.

We use the test signal as reference. Our method estimates the TDOA between m-th and n-th signals from l-th loudspeaker, $\tau^l_{m \to n}$, by two steps: (1) Calculate the cross correlations $R_{s,m_n}(\tau), R_{s,m_m}(\tau)$ between each channel and the test signal:

$$R_{s,m_n}(\tau) = E[s(t) m_n(t+\tau)] \tag{3}$$

$$R_{s,m_m}(\tau) = E[s(t) m_m(t+\tau)] \tag{4}$$

(2) Subtract the maximums of the cross correlations of m-th and n-th recorded signals to estimate TDOA:

$$\tau^l_{m \to n} = \arg\max_\tau R_{s,m_m}(\tau) - \arg\max_\tau R_{s,m_n}(\tau) \tag{5}$$

Since the test signal contains no noise signal, we can estimate the TDOA more accurate than the method using the cross correlation coefficient between recorded signals.

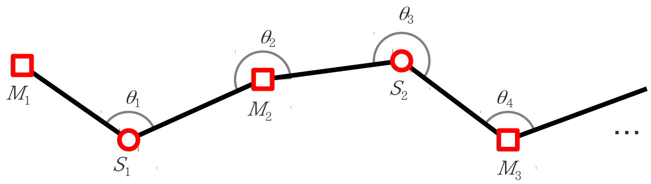


Fig. 3. Piecewise linear model of the hose-shaped robot posture.

C. Formulation of the robot shape

We represent the posture of the hose-shaped robot using a piecewise linear model. Figure 3 shows our expression of the robot posture. The microphones and loudspeakers are connected with a segment of $L/2$ length.

Since the length of each segment are a constant, the joint angles of the neighboring segments represent the whole posture of the hose-shaped robot. Therefore, the state vector consists of only joint angles:

$$\boldsymbol{x}_k = [\theta_0, \theta_1, \cdots \theta_{2N-3}]^T \tag{6}$$

The coordinates of the microphone and the loudspeaker are calculated from the first ones, $\boldsymbol{M}_1$ and $\boldsymbol{S}_1$, recursively:

$$\boldsymbol{M}_m = \boldsymbol{S}_{m-1} + l[\cos\theta'_{2(m-1)}, \sin\theta'_{2(m-1)}]^T \tag{7}$$

$$\boldsymbol{S}_l = \boldsymbol{M}_l + l[\cos\theta'_{2(l-1)}, \sin\theta'_{2(l-1)}]^T \tag{8}$$

$$\theta'_n = n\pi - \sum_{k=1}^{n} \theta_k \tag{9}$$

1) State update equation: During each test signal emission, we assumed that the posture of the hose-shaped robot is stable, the state vector performs random walk for each k:

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \boldsymbol{u}_k \tag{10}$$

where u_k denotes the Gaussian distribution represented by

$$\boldsymbol{u}_k = [\mathcal{N}(0, \sigma_u^2), \cdots, \mathcal{N}(0, \sigma_u^2)]^T \in \mathbb{R}^{2N-3} \tag{11}$$

2) Measurement equation: The TDOA between $\boldsymbol{M}_m$ and $\boldsymbol{M}_n$ from $\boldsymbol{S}_l$ is obtained from the Euclidean distance between the microphones and the loudspeaker:

$$\tau^l_{m\to n} = \frac{||\boldsymbol{M}_n - \boldsymbol{S}_l||}{c} - \frac{||\boldsymbol{M}_m - \boldsymbol{S}_l||}{c} \tag{12}$$

The measurement vector is a set of the TDOA between $\boldsymbol{M}_l$ and other microphones:

$$\boldsymbol{z}_k = \boldsymbol{h}(x) + \boldsymbol{v}_k \tag{13}$$

$$\boldsymbol{h}(x) = [\tau^l_{l\to 1}, \cdots, \tau^l_{l\to N}]^T \in \mathbb{R}^{N-2} \tag{14}$$

where v_k denotes the measurement noise represented by

$$\boldsymbol{v}_k = [\mathcal{N}(0, \sigma_v^2), \cdots, \mathcal{N}(0, \sigma_v^2)]^T \in \mathbb{R}^{N-2} \tag{15}$$

Since the distance $|\boldsymbol{M}_l - \boldsymbol{S}_l|$ and $|\boldsymbol{M}_{l+1} - \boldsymbol{S}_l|$ are the same value in this model, $\tau^l_{l\to l}$ and $\tau^l_{l\to l+1}$ are constant value in this model. Thus, they are not included in this measurement.

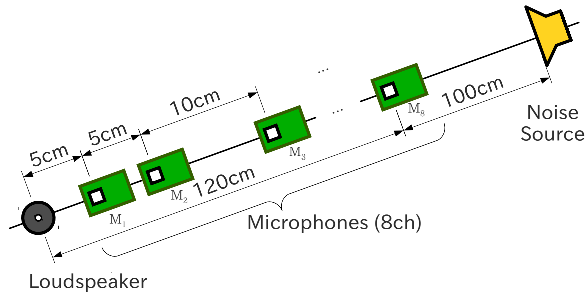


Fig. 4. Location of microphones and sources of our TDOA estimation.

D. Extended Kalman Filter

We built an EKF [14] based online microphone array localization. The EKF consists of the two steps: prediction and correction. It estimates posteriori state by correcting the current state estimates from a new measurement.

1) Prediction step: In the prediction step, we project the state and error covariance estimates from the previous step. From Eq. 10, the current time state estimates $\boldsymbol{x}_k^-$ and error covariance estimates $\boldsymbol{P}_k^-$ are updated:

$$\boldsymbol{x}_k^- = \boldsymbol{x}_{k-1} \tag{16}$$

$$\boldsymbol{P}_k^- = \boldsymbol{P}_{k-1} + \boldsymbol{Q}_{k-1} \tag{17}$$

where $\boldsymbol{Q}_{k-1}$ denotes the covariance matrix defined by

$$\boldsymbol{Q}_{k-1} = \mathrm{diag}(\sigma_u^2, \cdots, \sigma_u^2) \in \mathbb{R}^{2N-3\times 2N-3} \tag{18}$$

2) Correction step: In the correction step, we generate the Kalman gain $\boldsymbol{K}_k$ to minimize posteriori error covariance:

$$\boldsymbol{K}_k = \boldsymbol{P}_k^- \boldsymbol{H}^T (\boldsymbol{H}_k \boldsymbol{P}_k^- \boldsymbol{H}_k^T + \boldsymbol{R}_k)^{-1} \tag{19}$$

Then, we compute the posteriori state $\hat{\boldsymbol{x}}_k$ and error covariance $\boldsymbol{P}_k$ with the measurement $\boldsymbol{z}_k$:

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^- + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_k^-)) \tag{20}$$

$$\boldsymbol{P}_k = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\boldsymbol{P}_k^- \tag{21}$$

where $\boldsymbol{H}_k$ and $\boldsymbol{R}_k$ denote Jacobian of $\boldsymbol{h}(\boldsymbol{x})$ and the covariance matrix, respectively:

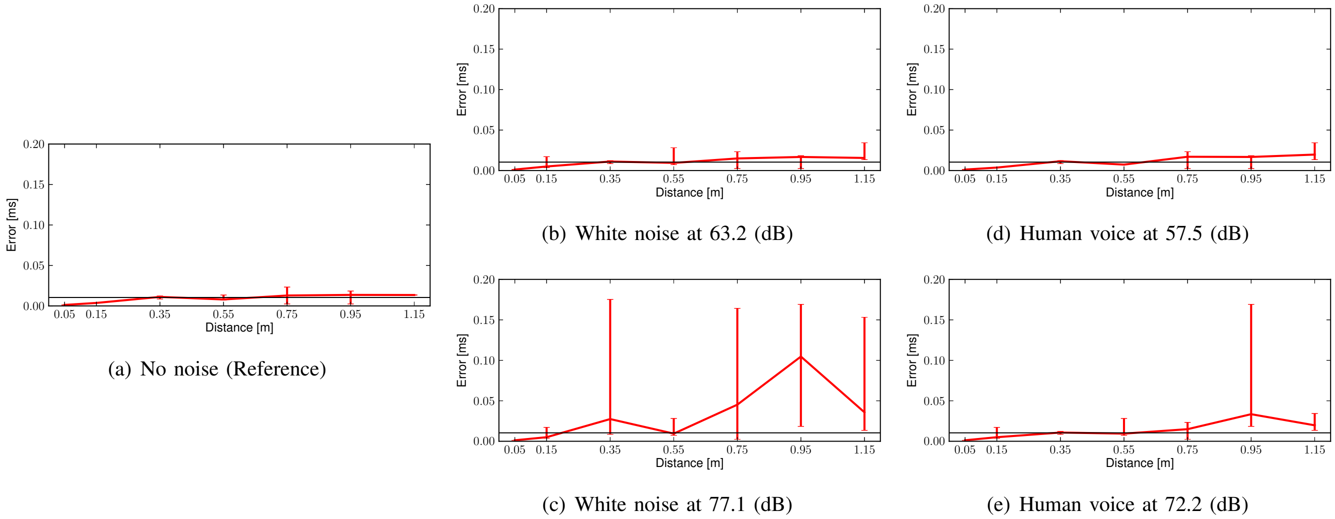
$$\boldsymbol{R}_k = \mathrm{diag}(\sigma_v^2, \cdots, \sigma_v^2) \in \mathbb{R}^{N-2\times N-2} \tag{22}$$

V. Evaluation

We evaluated our TDOA estimation method under two kinds of noise: white noise and human voice in Section V.A, and posture estimation with three shapes in Section V.B.

A. Evaluation 1: TDOA estimation

We evaluated our TDOA estimation method under the setting of white noise and human voice as directional noise.



(a) No noise (Reference)

(b) White noise at 63.2 (dB)

(c) White noise at 77.1 (dB)

(d) Human voice at 57.5 (dB)

(e) Human voice at 72.2 (dB)

Fig. 5. The errors of our TDOA estimation for various distance and two noises.

TABLE II

AVERAGE AND STANDARD DEVIATION OF EVALUATION 1.

Type of noise	Amplitude (dB SPL)	Mean (ms)	Standard deviation (ms)
No noise		0.00641	0.00938
White noise	63.2	0.00849	0.0101
White noise	77.1	−0.0159	0.0611
Human voice	57.5	0.00909	0.0106
Human voice	72.2	0.00641	0.0238

1) Settings: The evaluation was conducted in an anechoic chamber. Figure 4 shows the location of the loudspeaker, microphones and noise source. We put a small loudspeaker on the left side and a loudspeaker (MS101-III by YAMAHA) as the noise source on the right side. We put microphones spaced at 5, 10, 20, 40, 60, 80, 100, and 120 cm from the small loudspeaker.

The test signal was a sweep signal from 4 kHz to 8 kHz, 200 ms. The amplitude of the loudspeaker was calibrated so that its sound pressure level (SPL) was 61.0 dB when it plays a 4 kHz pure tone. We set the SPL of the white noise to 63.2 dB/77.1 dB and the human voice to 57.5 dB/72.2 dB. The human voice was the speech of 1st sentence of phonetically balanced sentences (ASJ-JNAS[15]). The sound pressure level of the noise source was measured from 1 m away from the loudspeaker by NL-21 by Rion Co. Ltd. with A weighted mode.

We recorded the acoustic signal sampled with 8 channels, 48 kHz, and 24 bits using a multichannel A/D converter called RASP24 developed by Systems In Frontier Corp. RASP24 can perform 8-channel synchronized recording. The test signal was sounded 10 times every 400 ms. We split the recorded signal into 400 ms to include one entire test signal, estimated the TDOA between M_1 and other microphones at each split signal, and calculated the error of the TDOA estimation.

2) Results: Figure 5 shows the results of the TDOA estimation with no noise, white noise, and human voice, respectively. Table II shows the average and standard deviation.

In our methods, the lower limit of the error is half of the sampling interval (the line of 0.0104 ms in the figures). In no noise situation, the error is lower than the limit in average. The maximum errors of the white noise at 63.2 dB and the human voice at 57.5 dB are smaller than 0.035 ms which is nearly 3.5 times the limit. Moreover, the average error are smaller than the lower limit. We can say that in these situations, our method successfully estimates the TDOA.

The errors of the white noise at 77.1 dB and the human voice at 72.2 dB are larger than the others according to Fig.5(c), 5(e). Since the cross correlation of the sweep signal and noise signal is invariably not zero, our TDOA estimation rises the error when the power of noise signal was too loud. This problem will be solved by employing noise suppression techniques.

B. Evaluation 2: EKF-based posture estimation

In this Section, we evaluated the posture estimation error with simulated data of three shapes: S-curve, circle, and straight.

1) Settings: The measurement vector was simulated by the microphone and loudspeaker locations of the correct posture and added measurement noise following a Gaussian distribution with standard deviation of no noise: $\sigma_v = 0.00938$ ms, which was evaluated in the anechoic chamber. The initial state was given by adding the uniform distribution (± 15 deg) to the correct posture state. The state value was restricted to between 120 deg and 240 deg considering the limitation of hose bending. We performed 100 evaluations for each shape and calculated the mean error of the microphone locations at each step.

2) Results: Figure 6 shows the estimation errors of microphone locations. The error bar indicates the minimum and maximum estimation errors and the polyline indicates the mean errors. Figure 7 shows the examples of the results for each posture and the correct postures.

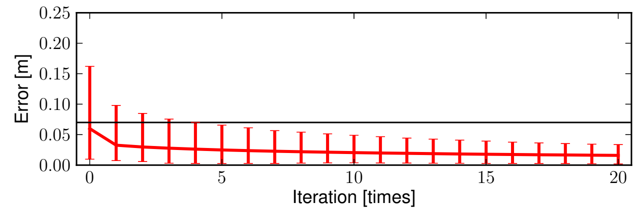
In the S-curve and circle shape, the errors become less than 10% of the hose length in about fifth and second iterations at worst case, respectively. This shows that our posture estimation works properly in these shapes.

On the other hand, in the straight shape, Fig. 6(c), the convergence speed was slower than the other two shapes. This is because the Jacobian converges to zero as the state values converge to 180 deg; in other words the posture converges straight. This shows that the first order approximation of EKF was not tolerable for the hose-shaped robot space model. For solving this problem, we will introduce a non-linear filter such as UKF or particle filter.

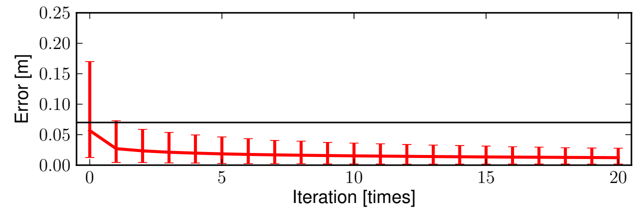
VI. Conclusion

This paper presented a posture estimation of the hose-shaped robot by localizing microphone positions with microphones and small loudspeakers equipped on the hose. Our method is built with a sweep signal and a piecewise linear model of the hose-shaped robot posture using an EKF. The main contribution is that our method demonstrates the posture estimation of the hose-shaped robot by sound. Experimental results show that our TDOA estimation can estimate properly and our posture estimation method has good potential.

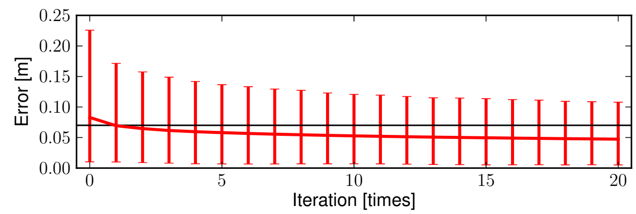
For enhancing navigational usability, we can estimate the posture of the hose-shaped robot more precisely by integrating the microphones with the conventional sensors, such as, video cameras, inertial sensors, and magnetic sensors. In addition, we can enhance mission usability because the base of sound source localization and separation was established, which is the microphone array localization for the hose-shaped robot. We will develop these functions for practical use.



(a) S-curve

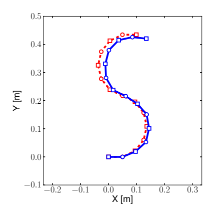


(b) Circle

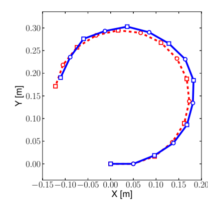


(c) Straight

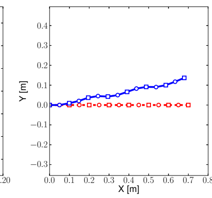
Fig. 6. The errors of the posture estimation for each iteration.



(a) S-curve



(b) Circle



(c) Straight

Fig. 7. The examples of the posture estimation results for each posture (blue) and the correct postures (red). Squares and circles indicate microphones and loudspeakers, respectively.

Acknowledgments

We are grateful to Prof. Satoshi Tadokoro, Prof. Masashi Konyo, and Junnichi Fukuda of Tohoku University for giving a material of the hose-shaped robot and helpful advice. This study is partially supported by a Grant-in-Aid for Scientific Research (S) (No.24220006).

References

[1] R. R. Murphy, "Navigational and Mission Usability in Rescue Robots," *Robotics Society of Japan*, vol. 28, no. 2, pp. 142–146, 2010.

[2] T. Mizumoto *et al.*, "Design and Implementation of Selectable Sound Separation on Texai Telepresence System using HARK," *IEEE ICRA 2011*, pp. 2130–2137, 2011.

[3] K. Nakadai *et al.*, "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, 2011.

[4] A. Kitagawa *et al.*, "Development of Small Diameter Active Hose-II For Search and Life-prolongation of Victims under Debris," *Robotics and Mechatoronics*, vol. 15, no. 5, pp. 474–481, 2003.

[5] T. Kamegawa *et al.*, "Development of The Snake-like Rescue Robot KOHGA," *IEEE ICRA 2004*, pp. 5081–5086, 2004.

[6] M. Arai *et al.*, "Development of "Souryu-IV" and "Souryu-V:" Serially Connected Crawler Vehicles for in-Rubble Searching Operations," *Field Robotics*, vol. 25, no. 1–2, pp. 31–65, 2008.

[7] K. Hatazaki *et al.*, "Active Scope Camera for Urban Search and Rescue," *IEEE/RSJ IROS 2007*, pp. 2596–2602, 2007.

[8] S. Tadokoro *et al.*, "Application of Active Scope Camera to Forensic Investigation of Construction Accident," *IEEE ARSO 2009*, pp. 48–59, 2009.

[9] H. Namari *et al.*, "Tube-type Active Scope Camera with High Mobility and Practical Functionality," *IEEE/RSJ IROS 2012*, pp. 2679–2686j, 2012.

[10] M. Ishikura *et al.*, "Vision-based localization using Active Scope Camera - Accuracy Evaluation for Structure from Motion in Disaster Environment," *IEEE/SICE SII 2010*, pp. 25–30, 2010.

[11] M. Ishikura, "Shape Estimation of Flexible Cable," *IEEE/RSJ IROS 2012*, pp. 2539–2546, 2012.

[12] N. Ono *et al.*, "Blind alignment of asynchronously recorded signals for distributed microphone array," *IEEE WASPAA*, pp. 161–164, 2009.

[13] H. Miura *et al.*, "SLAM-based online calibration of asynchronous microphone array for robot audition," *IEE/RSJ IROS 2011*, pp. 524–529, 2011.

[14] G. Welch *et al.*, "An introduction to the Kalman filter," TR95-041, Department of Computer Science, University of North Carolina, Chapel Hill, 2000.

[15] K. Itou *et al.*, "The Design of the Newspaper-Based Japanese Large Vocabulary Continuous Speech Recognition Corpus," *ICSCP 1998*, pp. 3261–3264, 1998.