A Sound-based Online Method for Estimating the Time-Varying Posture of a Hose-shaped Robot

Yoshiaki Bando*, Katsutoshi Itoyama*, Masashi Konyo†, Satoshi Tadokoro†,
Kazuhiro Nakadai‡, Kazuyoshi Yoshii* and Hiroshi G. Okuno§
*Graduate School of Informatics, Kyoto Univ., Kyoto, 606-8501, Japan, {yoshiaki, itoyama, yoshii}@kuis.kyoto-u.ac.jp
†Graduate School of Information Science, Tohoku Univ., Miyagi, 980-8579, Japan, {konyo, tadokoro}@rm.is.tohoku.ac.jp
‡Honda Research Institute Japan Co., Ltd., Saitama, 351-0114, Japan, nakadai@jp.honda-ri.com
§Graduate Program for Embodiment Informatics, Waseda Univ., Tokyo 169-0072, Japan, okuno@aoni.waseda.jp

***Abstract*—This paper presents an online method that can accurately estimate the time-varying posture of a moving hose-shaped robot having multiple microphones and loudspeakers. Sound-based posture estimation has been considered to be promising for circumventing the cumulative error problem of conventional integral-type methods using differential information obtained by inertial sensors. Our robot emits a reference signal from a loudspeaker one by one and estimates its posture by measuring the time differences of arrival (TDOAs) at the microphones. To accurately estimate the posture of the robot (the relative positions of the microphones and loudspeakers) even when the robot moves, we propose a novel state-space model that represents the dynamics of not only the posture itself but also its change rate in the state space. This model is used for predicting the current posture by using an unscented Kalman filter. The experiments using a 3 m moving hose-shaped robot with eight microphones and seven loudspeakers showed that our method achieved less than 20 cm error at the tip position even after the robot moved over a long time, whereas the estimation error obtained by a conventional integral-type method increased monotonically over time.**

I. Introduction

Hose-shaped robots [1]–[3] are one of the most useful types of rescue robots that can be used for probing buried victims in a disaster environment where humans or animals cannot work [4]–[7]. Those robots have thin and long bodies and can penetrate into narrow gaps in the rubble of collapsed buildings. A remote operator steers a hose-shaped robot to the target location by using its locomotion mechanism. Active Hose-II [1], for example, has small powered wheels to move forward. Active Scope Camera [2], [3] has a body covered with cilia and can move forward by vibrating the cilia. It was used for real search-and-rescue in Jacksonville, Florida, USA in 2008 [8].

To control a hose-shaped robot that flexibly changes its posture (shape) over time in an unseen environment, it is necessary to estimate the time-varying posture of the moving robot. Ishikura *et al.* [9], for example, proposed an inertial-sensor-based method that can estimate the posture by integrating the acceleration and angular-velocity information obtained from accelerometers and gyro sensors installed on the robot. Such integral-type methods based on the posture change rate, however, cannot work over a long time because the estimation error is gradually accumulated. Although non-integral-type methods based on information obtained by GPS and strain gauges can accurately track the posture independently of the past history [10]–[12], those methods can neither be used indoors nor be used for a robot with a long body [13].

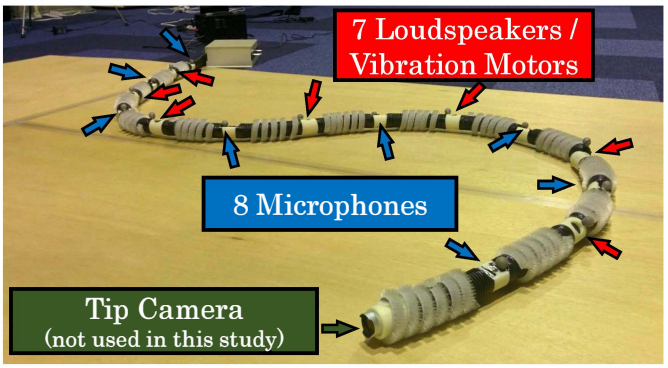


Fig. 1. A prototype hose-shaped robot with a driving mechanism. Microphones and loudspeakers are used to estimate their relative positions.

(a) Microphone (b) Loudspeaker

Fig. 2. Microphone and loudspeaker on the prototype hose-shaped robot.

Sound-based posture estimation has recently been considered to be a promising non-integral-type approach. A hose-shaped robot having multiple microphones and loudspeakers, for example, has been developed [14]. This robot emits a reference signal from a loudspeaker one by one and estimates its posture by measuring the time differences of arrival (TDOAs) at the microphones. Since those TDOAs depend only on the current relative positions of the microphones and loudspeakers, the cumulative error problem can be avoided. The sound-based approach can be used in a closed space allowing sound propagation, whereas the accurate GPS-based approach can be used only outdoors for receiving signals from the satellites. This indicates that sound-based posture estimation is complementary to inertial-sensor-based and GPS-based posture estimation. The robot audition mechanism is useful for sound source localization and separation [15], [16] (*e.g.*, search for victims by voice) as well as posture estimation.

A major requirement of posture estimation is that the robot posture should be continuously presented to an operator in real time. Ono *et al.* [17] for example, proposed a method based on auxiliary functions for performing simultaneous localization of

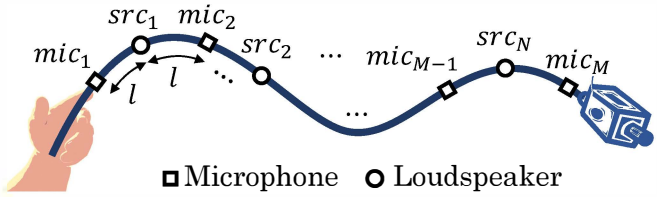


Fig. 3. Microphone and loudspeaker arrangements.

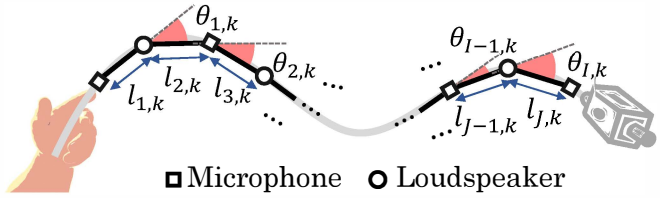


Fig. 4. Serially-connected link model of robot posture.

asynchronous microphones and multiple sound sources. Since this method is intended for offline use and assumes that the sound sources and microphones are stable, it cannot be used for posture estimation of the moving robot. Miura *et al.* [18] proposed a sound-based method of simultaneous localization and mapping (SLAM). Although this method can be used in an online manner, it assumes that there is a single moving sound source with known dynamics.

In this paper we propose a new sound-based online method that can accurately estimate the time-varying posture of a moving hose-shaped robot. To achieve this, we formulate a state-space model that represents the dynamics of not only the posture itself but also its change rate in the state space. Our model has two distinct characteristics. First, to use our method in an online manner, the current posture of the robot is predicted from the previous posture by using an unscented Kalman filter (UKF) [19]. Second, our model assumes that the relative positions of the microphones and loudspeakers can change over time under a constraint that the microphones and loudspeakers are serially linked in a specified order. The effectiveness of our proposed method was evaluated using a prototype hose-shaped robot with a driving mechanism, as shown in Fig. 1.

The remainder of the paper is organized as follows. Section II presents the sound-based online posture estimation method. Section III shows and discusses the experimental results for a prototype hose-shaped robot. Finally, Section IV summarizes the key findings and mentions future work.

II. Sound-based Online Posture Estimation

This section describes our proposed method of sound-based online posture estimation. The posture of a hose-shaped robot is estimated according to the following three steps: 1) generate a reference signal from each loudspeaker, one by one, 2) estimate the TDOAs of the reference signal at the microphones, and 3) estimate the relative positions of the microphone and loudspeakers from the estimated TDOAs.

A. Problem Statement

A hose-shaped robot we use has microphones and loudspeakers installed alternately at a regular interval l, as shown in Fig. 3. We denote the microphones and loudspeakers as mic_m $(m = 1, \cdots, M)$ and src_n $(n = 1, \cdots, N)$, respectively, where $N = M - 1$. We define k as the measurement index and the microphone and loudspeaker positions as $\boldsymbol{x}^{\mathrm{mic}}_{m,k}$ and $\boldsymbol{x}^{\mathrm{src}}_{n,k}$, respectively. In this paper, we assume that the microphones and loudspeakers are on a two-dimensional surface. The other notations are summarized in Table I.

The problem statement for a sound-based posture estimation is defined as follows:

TABLE I. Definition of Mathematical Symbols

Symbol	Meaning
M	Number of microphones
N	Number of loudspeakers ($N = M - 1$)
C	Speed of sound
t	Time
k	Measurement index
ω	Frequency
mic_m	i-th microphone ($1 \leq m \leq M$)
src_n	j-th loudspeaker ($1 \leq n \leq N$)
$\boldsymbol{x}^{\mathrm{mic}}_{m,k}$	Position of $\mathrm{mic}_m \in \mathbb{R}^2$
$\boldsymbol{x}^{\mathrm{src}}_{n,k}$	Position of $\mathrm{src}_n \in \mathbb{R}^2$
$\boldsymbol{z}_k(t)$	Input audio recording at k-th measurement $\in \mathbb{R}^M$
$s(t)$	Reference signal $\in \mathbb{R}$
$\boldsymbol{\xi}_k$	State variable at k-th measurement $\in \mathbb{R}^{4(M+N)-6}$
$\boldsymbol{\zeta}_k$	Posture at k-th measurement $\in \mathbb{R}^{2(M+N)-3}$
$\theta_{\bullet,k}$	Joint angle ($\theta_{\bullet,k} \in \mathbb{R}$, $1 \leq a \leq N + M - 2$)
$l_{\blacklozenge,k}$	Link length ($l_{\blacklozenge,k} \in \mathbb{R}$, $1 \leq b \leq M + N - 1$)
$\dot{\boldsymbol{\zeta}}_k$	Posture change rate at k-th measurement $\in \mathbb{R}^{2(M+N)-3}$
$\boldsymbol{y}_k$	k-th measurement vector $\in \mathbb{R}^{M-1}$
$\tau^n_{m_1,m_2,k}$	TDOA between mic_{m_1} and mic_{m_2} for a reference signal generated by $\mathrm{src}_n \in \mathbb{R}$

Input:
Synchronized M-channel audio signals $\boldsymbol{z}_k(t)$ obtained by recording a reference signal $s(t)$ with M microphones.
Output:
The relative positions of each microphone $\boldsymbol{x}^{\mathrm{mic}}_{m,k}$ and each loudspeaker $\boldsymbol{x}^{\mathrm{src}}_{n,k}$.

The input data are used for calculating the TDOA of the reference signal at each microphone. Since the TDOA represents the relationship between the microphone and loudspeaker, the output is the *relative* positions of the microphones and loudspeakers. We therefore assume that $\boldsymbol{x}^{\mathrm{mic}}_{1,k}$ and $\boldsymbol{x}^{\mathrm{src}}_{1,k}$ are known without loss of generality.

B. State-Space Model of Robot Posture

Our method estimates the posture of a moving robot by using the TDOAs calculated from the input data. More specifically, we formulate a nonlinear state-space model that associates a state space representing the posture dynamics with an observation space representing the TDOA. The point estimate of the current posture is obtained by using an UKF.

The robot posture is modeled as a serially-connected link model, as shown in Fig. 4. The posture at the k-th measurement, $\boldsymbol{\zeta}_k$, is defined as

$$\boldsymbol{\zeta}_k = [\theta_{1,k}, \cdots, \theta_{M+N-2,k}, l_{1,k}, \cdots, l_{M+N-1,k}], \quad (1)$$

where $\theta_{a,k}$ $(1 \leq a \leq M + N - 2)$ is a link angle and $l_{b,k}$ $(1 \leq b \leq M + N - 1)$ is a link length. To deal with a moving robot, we estimate not only posture $\boldsymbol{\zeta}_k$ but also its change rate, $\dot{\boldsymbol{\zeta}}_k$. The state-space vector $\boldsymbol{\xi}_k$ is given by

$$\boldsymbol{\xi}_k = [\boldsymbol{\zeta}_k, \dot{\boldsymbol{\zeta}}_k]^{\mathrm{T}} \in \mathbb{R}^L, \quad (2)$$

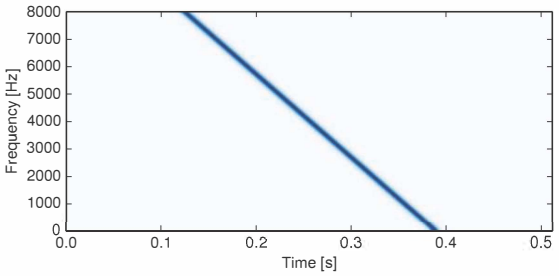


Fig. 5. TSP signal with length of 8192 samples at 16 kHz.

where $L = 4M + 4N - 6$ is the dimension of the state space.

The relative positions of the microphones and loudspeakers on the robot, $\boldsymbol{x}^{\mathrm{mic}}_{m,k}$ and $\boldsymbol{x}^{\mathrm{src}}_{n,k}$, can be calculated recursively from the known positions $\boldsymbol{x}^{\mathrm{mic}}_{1,k}$ and $\boldsymbol{x}^{\mathrm{src}}_{1,k}$. Suppose that $\boldsymbol{x}^{*}_{i,k}$ is the i-th member of $[\boldsymbol{x}^{\mathrm{mic}}_{1,k}, \boldsymbol{x}^{\mathrm{src}}_{1,k}, \cdots, \boldsymbol{x}^{\mathrm{mic}}_{M-1,k}, \boldsymbol{x}^{\mathrm{src}}_{N,k}, \boldsymbol{x}^{\mathrm{mic}}_{M,k}]$, each position is given by

$$\boldsymbol{x}^{*}_{i,k} = \boldsymbol{x}^{*}_{i-1,k} + l_{i,k} \times \left[\cos\left(\sum_{a=1}^{i-1}\theta_{a,k}\right), \sin\left(\sum_{a=1}^{i-1}\theta_{a,k}\right)\right]^{\mathrm{T}}.$$

1) Measurement Model: A measurement model $p(\boldsymbol{y}_k|\boldsymbol{\xi}_k)$ is formulated using TDOA $\tau^{n}_{m_1,m_2,k}(\boldsymbol{\xi}_k)$ between mic_{m_1} and mic_{m_2} for the reference signal generated from src_n as follows:

$$p(\boldsymbol{y}_k|\boldsymbol{\xi}_k) = \mathcal{N}(\boldsymbol{y}_k|\boldsymbol{T}(\boldsymbol{\xi}_k), \boldsymbol{R}_k), \tag{3}$$

$$\boldsymbol{T}(\boldsymbol{\xi}_k) = [\tau^{n}_{n,1,k}(\boldsymbol{\xi}_k), \cdots, \tau^{n}_{n,n-1,k}(\boldsymbol{\xi}_k), \tau^{n}_{n,n+1,k}(\boldsymbol{\xi}_k), \cdots, \tau^{n}_{n,M,k}(\boldsymbol{\xi}_k)]^{\mathrm{T}} \tag{4}$$

TDOA $\tau^{n}_{m_1,m_2,k}$ is calculated by using the distances between the two microphone and the loudspeaker as follows:

$$\tau^{n}_{m_1,m_2,k}(\boldsymbol{\xi}_k) = \frac{|\boldsymbol{x}^{\mathrm{mic}}_{m_2,k} - \boldsymbol{x}^{\mathrm{src}}_{n,k}| - |\boldsymbol{x}^{\mathrm{mic}}_{m_1,k} - \boldsymbol{x}^{\mathrm{src}}_{n,k}|}{C}, \tag{5}$$

where C represents the speed of sound. We assume that C is 340 m/s in this paper.

2) State Update Model: A state update model $p(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ is based on two concepts: a) posture dynamics and b) posture constraint. The posture dynamics $q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ represents how likely the previous posture $\boldsymbol{\zeta}_{k-1}$ is to change to the current posture $\boldsymbol{\zeta}_k$ with a change rate $\dot{\boldsymbol{\zeta}}_{k-1}$ as follows:

$$q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) = \mathcal{N}(\boldsymbol{\xi}_k|[\boldsymbol{\zeta}_{k-1} + \dot{\boldsymbol{\zeta}}_{k-1}, \dot{\boldsymbol{\zeta}}_{k-1}]^{\mathrm{T}}, \boldsymbol{Q}_k), \tag{6}$$

where $\boldsymbol{Q}_k \in \mathbb{R}^{L\times L}$ is the covariance matrix of the process noise. The posture constraint $r(\boldsymbol{\xi}_k)$, on the other hand, is modeled as a Gaussian distribution:

$$r(\boldsymbol{\xi}_k) = \mathcal{N}(\boldsymbol{\xi}_k|\boldsymbol{\xi}, \boldsymbol{P}), \tag{7}$$

where $\boldsymbol{\xi} \in \mathbb{R}^{L}$ and $\boldsymbol{P} \in \mathbb{R}^{L\times L}$ are the mean and covariance matrix of the feasible posture.

We integrate these two distributions for the state update model $p(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ on the basis of the product of experts [20]:

$$p(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) = \frac{1}{A} q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) r(\boldsymbol{\xi}_k), \tag{8}$$

where $A = \int q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) r(\boldsymbol{\xi}_k) d\boldsymbol{\xi}_k$ is the normalization factor.

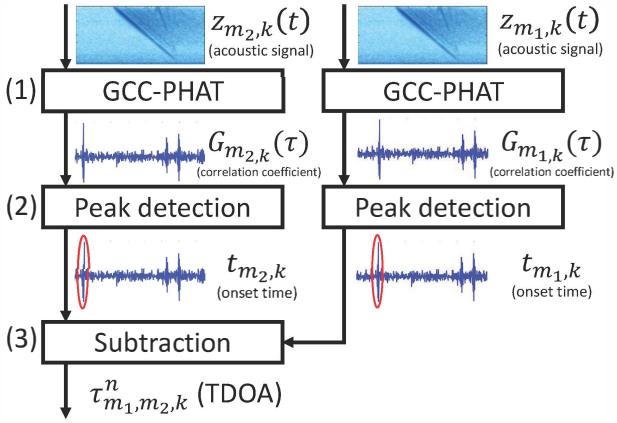


Fig. 6. Overview of TDOA estimation.

3) Estimation Algorithm: The robot posture $\boldsymbol{\zeta}_k$ is estimated from $\boldsymbol{y}_{1:k}$ in an online manner by using an UKF [19] assuming that the posterior distribution of the state variable $\boldsymbol{\xi}_k$ follows a Gaussian distribution. The UKF approximates the posterior distribution $p(\boldsymbol{\xi}_k|\boldsymbol{y}_{1:k})$ from the likelihood $p(\boldsymbol{y}_k|\boldsymbol{\xi}_k)$ and prior $p(\boldsymbol{\xi}_k|\boldsymbol{y}_{1:k-1})$ using unscented transform. The prior distribution $p(\boldsymbol{\xi}_k|\boldsymbol{y}_{1:k-1})$ is given by $\int p(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1}) p(\boldsymbol{\xi}_{k-1}|\boldsymbol{y}_{1:k-1}) d\boldsymbol{\xi}_{k-1}$ using unscented transform.

In our state-space model, we can simplify the calculation of the prior distribution $p(\boldsymbol{\xi}_k|\boldsymbol{y}_{1:k-1})$. Since the $q(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k-1})$ is a linear transformation of $\boldsymbol{\xi}_{k-1}$ (Eq. 6) and the $r(\boldsymbol{\xi}_k)$ is defined as a Gaussian distribution (Eq. 7), the state update model can be written as a linear model. We can therefore calculate the prior distribution $p(\boldsymbol{\xi}_k|\boldsymbol{y}_{1:k-1})$ without unscented transform as follows:

$$p(\boldsymbol{\xi}_k|\boldsymbol{y}_{1:k-1}) = \mathcal{N}(\boldsymbol{\xi}_k|\boldsymbol{\xi}^{-}_k, \boldsymbol{P}^{-}_k), \tag{9}$$

$$\boldsymbol{\xi}^{-}_k = \boldsymbol{P}^{-}_k((\boldsymbol{P}^{*}_k)^{-1}\hat{\boldsymbol{\xi}}_{k-1} + \boldsymbol{P}^{-1}\boldsymbol{\xi}), \tag{10}$$

$$\boldsymbol{P}^{-}_k = ((\boldsymbol{P}^{*}_k)^{-1} + \boldsymbol{P}^{-1})^{-1}, \tag{11}$$

$$\boldsymbol{P}^{*}_k = \boldsymbol{F}^{\mathrm{T}}\hat{\boldsymbol{P}}_{k-1}\boldsymbol{F} + \boldsymbol{P}_k, \tag{12}$$

where $\hat{\boldsymbol{\xi}}_{k-1}$ and $\hat{\boldsymbol{P}}_{k-1}$ are the mean vector and covariance matrix of the last posterior distribution $p(\boldsymbol{\xi}_{k-1}|\boldsymbol{y}_{1:k-1})$. This calculation is recursively performed over time.

C. Robust TDOA Estimation

To make TDOA estimation robust against motor noise, we use a time stretched pulse (TSP) [21] as a reference signal (Fig. 5). A TSP has a high signal-to-noise ratio and can be sent with large energy from a loudspeaker. Therefore, the reference signal can be easily distinguished from the motor noise. A TSP signal with a length of W samples is defined in the frequency domain as follows:

$$S(\omega) = \begin{cases} \exp(j2\pi\omega^2/W^2) & 0 \le \omega \le W/2 \\ S(W-\omega) & W/2 \le \omega \le W \end{cases}, \tag{13}$$

where $S(\omega)$ is the frequency spectrum of the reference signal $s(t)$ in the frequency domain and ω indicates a frequency. The reference signal $s(t)$ is obtained by the inverse discrete Fourier transform of $S(\omega)$.

As shown in Fig. 6, TDOA $\tau^{n}_{m_1,m_2,k}$ is estimated from the recorded signal $\boldsymbol{z}_k(t)$ as follows:

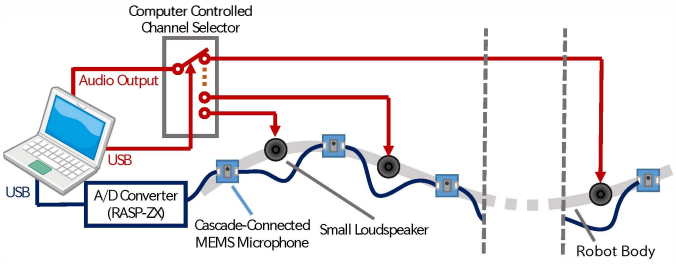


Fig. 7. System architecture of prototype hose-shaped robot.

1) Calculate the cross correlation coefficient $G_{m,k}(\tau)$ between each recorded signal $z_{m,k}(t)$ and the reference signal $s(t)$.
2) Calculate the onset times of the input signals, $t_{m_1,k}$ and $t_{m_2,k}$ by detecting the first peak of the correlation coefficient $G_{m_1,k}(\tau)$ and $G_{m_2,k}(\tau)$, respectively.
3) Calculate the TDOA $\tau^n_{m_1,m_2,k}$ by subtracting $t_{m_1,k}$ from $t_{m_2,k}$.

The cross correlation is calculated using the generalized cross correlation method with phase transform (GCC-PHAT) [22]. This method is robust against reverberation [23] because indoor environments, where the hose-shaped robots are to be used, occur reverberation.

III. Evaluation

This section reports the experiments that were conducted for evaluating the proposed method of online posture estimation using a prototype hose-shaped robot.

A. Experimental Conditions

Fig. 1 shows a prototype hose-shaped robot with a driving mechanism. The body was a corrugated tube with a diameter of 38 mm and a total length of 3 m. The driving mechanism was the same as that of a hose-shaped robot called tube-type Active Scope Camera [3]. More specifically, the whole surface of the robot was covered by cilia and the robot moved forward by vibrating the cilia using seven vibrating motors positioned at an interval of 40 cm. $M = 8$ microphones (Fig. 2(a)) and $N = 7$ loudspeakers (Fig. 2(b)) were positioned on the robot alternately, as shown in Fig. 1. The distance between the microphones at both ends was 2.8 m. We used a multichannel A/D converter with a sampling rate of 16 kHz and a quantization of 16 bit (RASP-ZX manufactured by Systems In Frontier Corp). The system architecture of our robot is shown in Fig. 7.

We compared our proposed method that can take into account the posture change rate with a conventional method that does not consider it. The initial shape of the robot was set to one of three postures: C-shape, S-shape, and straight. The experiment was conducted in an experimental room with a reverberation time (RT_{60}) of 800 ms (Fig. 8). The TSP reference signal had a length of 8192 samples (512 ms) at 16 kHz. The reference signal was recorded by the microphones using the HARK open source robot audition software [24]. To use UKF, we determined the initial state $\boldsymbol{\xi}_0 = [\boldsymbol{\zeta}_0, \dot{\boldsymbol{\zeta}}_0]$ in the following manner. The initial posture $\boldsymbol{\zeta}_0$ was sampled from a Gaussian distribution whose mean corresponds to the correct posture and

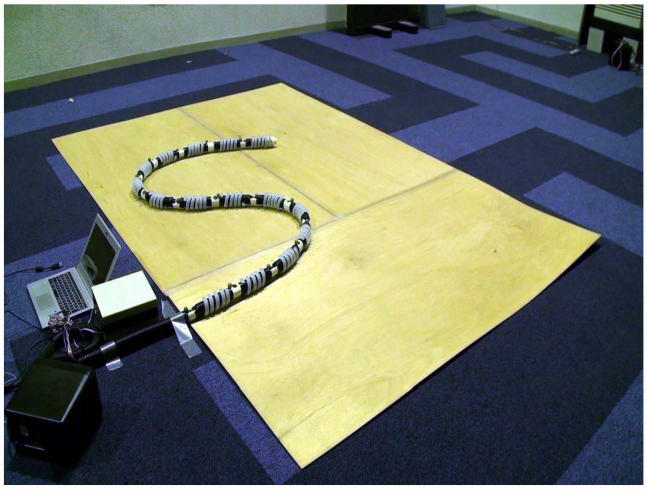


Fig. 8. Prototype hose-shaped robot placed on experimental room.

standard deviation was 15°. The initial change rate $\dot{\boldsymbol{\zeta}}_0$ was set to zero. The other parameters were determined experimentally.

The estimation algorithm was implemented using Python without multiprocessing. A standard laptop computer with an Intel Core i7-3517U CPU (2 cores, 1.9 GHz) and 4.0 GB of memory was used to estimate the TDOAs of the reference signal and the posture of the robot. The CPU time and elapsed time for 50 TDOA estimations (25.6 s) were 8.759 s and 8.843 s, respectively. Those for posture estimation were 2.679 s and 2.697 s, respectively. Therefore, the total computation time for an input signal of 25.6 s was 11.456 s.

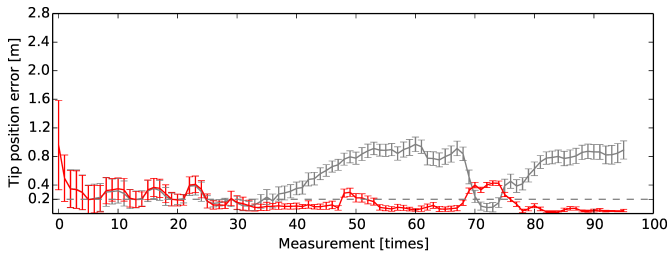
We evaluated the estimation error at the tip position. More specifically, the estimation error was calculated by measuring the difference between the estimated and correct positions of the microphone mic_8 that is the most distant from the reference points mic_1 and src_1. The correct positions were captured using a motion capture system (OptiTrack manufactured by NaturalPoint Inc.). The average estimation error was calculated over 32 different initial states.

B. Experimental Results

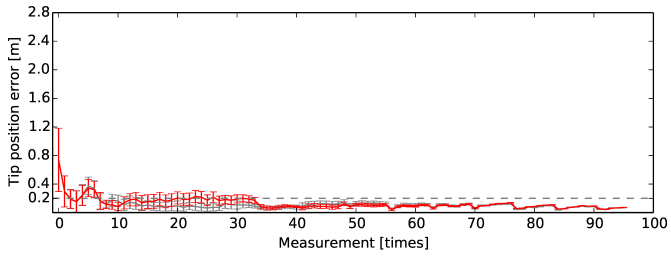
When the initial posture was set to the C-shape or S-shape, as shown in Figs. 9(a), 9(b), 10(a), and 10(b), the estimation errors were decreased over time and, as shown in Figs. 11 and 12, the estimated postures followed the moving robot postures accurately. Moreover, when the initial posture was set to the C-shape, the baseline method failed to follow the moving posture and the estimation error increased after the 30-th measurement. On the other hand, the proposed method successfully tracked the moving posture in real time.

The estimation errors, when the initial posture was set to the C-shape or S-shape, were almost under 0.2 m after the 40-th measurement. Ishikura *et al.* [9] reported that their inertial-sensor-based method achieved the estimation error about 0.2 m. Our method attained the similar performance to that of the inertial-sensor-based method.

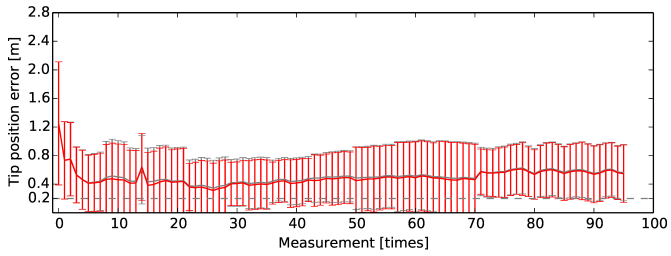
When the initial posture was straight, as shown in Figs. 9(c) and 10(c), on the other hand, the estimation error was larger than those obtained in the cases of the other initial postures.



(a) C-shape

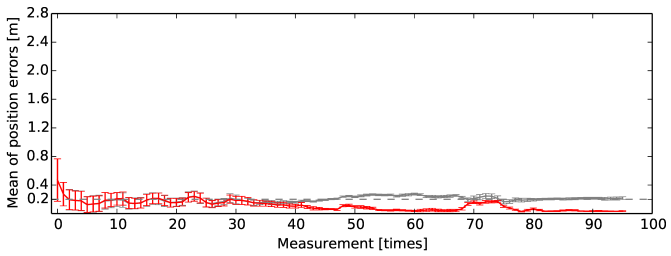


(b) S-shape

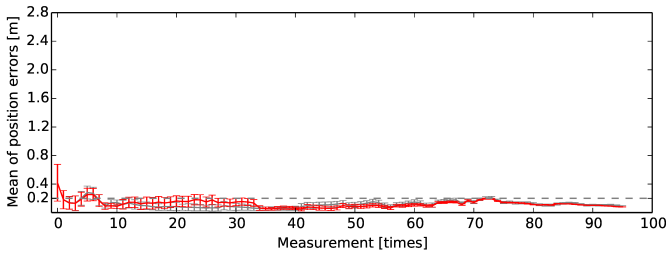


(c) Straight

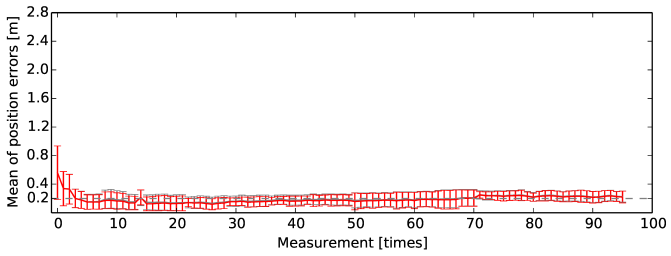
Fig. 9. The estimation errors obtained by the proposed and baseline methods for a moving prototype hose-shaped robot. The red line represents the proposed method, and the gray line represents the baseline method. The polyline and error bar indicate the mean and standard deviation, respectively.



(a) C-shape



(b) S-shape



(c) Straight

Fig. 10. The average estimation errors obtained by the proposed and baseline methods for a moving prototype hose-shaped robot. The red line represents the proposed method, and the gray line represents the baseline method. The polyline and error bar indicate the mean and standard deviation, respectively.

This is because of the mirror-symmetrical problem. Since the microphones and loudspeakers were installed on the robot in forming single row, we cannot distinguish between two postures which were mirror-symmetrical with respect to mic_1 and src_1. As shown in Figs. 13 and 14, the mirror-symmetrical postures were estimated.

A promising solution to this problem would be to use multimodal information, *i.e.*, integrate various types of information obtained from microphones, accelerometers, and gyro sensors. If a robot has those modalities, mirror-symmetrical postures can be distinguished by considering the posture change history and the robot can work in a closed and narrow space in which some modalities do not work. The mirror-symmetrical ambiguity could be handled with an unscented particle filter [25] that can maintain multiple possibilities about the posture of the robot at the same time.

IV. Conclusion

This paper presented an online method that can accurately estimate the time-varying posture of a moving hose-shaped robot having multiple microphones and loudspeakers. The experiments using a 3 m moving hose-shaped robot showed that our method successfully suppressed the estimation error under 20 cm at the tip position even after the robot moved over a long time, whereas the estimation error obtained by a conventional integral-type method increased monotonically over time. We found that our purely sound-based method often confuses mirror-symmetrical postures, depending on the initial value of the estimation.

To solve the mirror-symmetrical problem and improve the accuracy of posture estimation in a wide variety of realistic environments and situations, we plan to equip the robot with accelerometers and gyro sensors. The probabilistic state-space modeling enables us to integrate various types of information obtained from multi-modal sensors in a principled way. To evaluate the effectiveness of the proposed robot from the viewpoint of search-and-rescue, we plan to conduct more comprehensive experiments in a simulated disaster environment (*e.g.*, narrow and closed space).

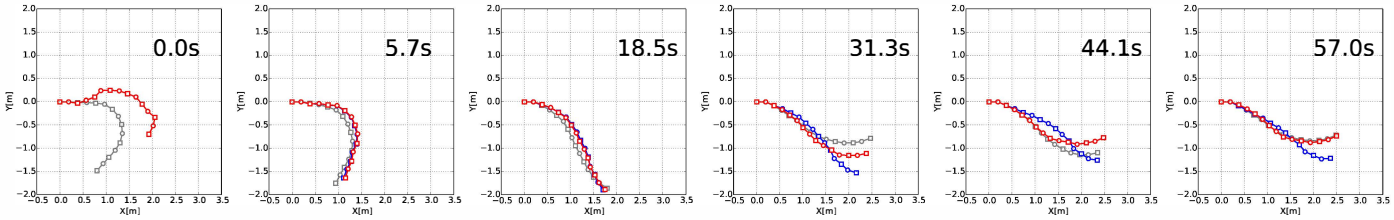
Acknowledgment

This study was partially supported by JSPS KAKENHI 24220006.

References

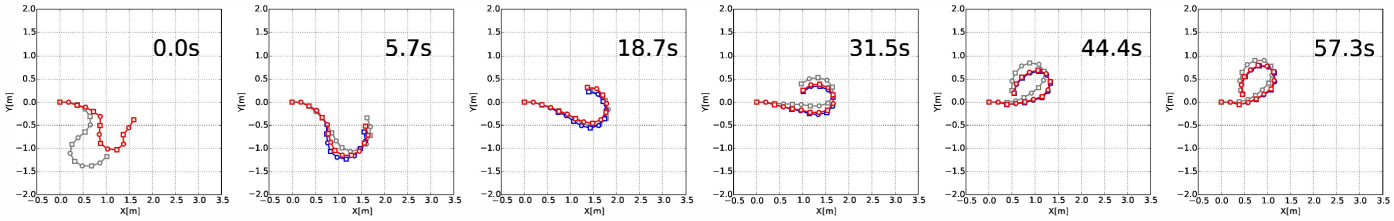
[1] A. Kitagawa *et al.*, "Development of small diameter Active Hose-II for search and life-prolongation of victims under debris," *Journal of Robotics and Mech.*, vol. 15, no. 5, pp. 474–481, 2003.

[2] K. Hatazaki, Konyo *et al.*, "Active scope camera for urban search and rescue," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007, pp. 2596–2602.



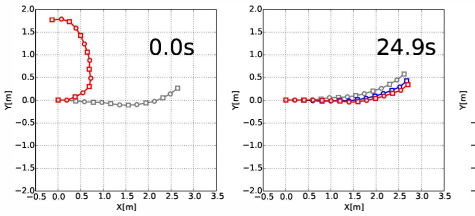
(a) Initial (b) 10-th measurement (c) 30-th measurement (d) 50-th measurement (e) 70-th measurement (f) 90-th measurement

Fig. 11. Estimation results when the initial posture was set to the C-shape. The red and blue lines indicate the postures estimated by the proposed and baseline methods, respectively. The gray line illustrates the correct posture.



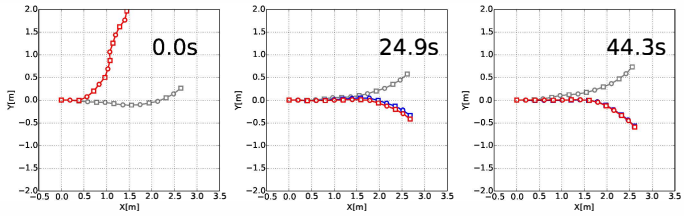
(a) Initial (b) 10-th measurement (c) 30-th measurement (d) 50-th measurement (e) 70-th measurement (f) 90-th measurement

Fig. 12. Estimation results when the initial posture was set to the S-shape. The red and blue lines indicate the postures estimated by the proposed and baseline methods, respectively. The gray line illustrates the correct posture.



(a) Initial (b) 40-th measurement (c) 70-th measurement

Fig. 13. Estimation results when the initial posture was set to straight. The red and blue lines indicate the postures estimated by the proposed and baseline methods, respectively. The gray line illustrates the correct posture.



(a) Initial (b) 40-th measurement (c) 70-th measurement

Fig. 14. Estimation results when the initial posture was set to straight. The red and blue lines indicate the postures estimated by the proposed and baseline methods, respectively. The gray line illustrates the correct posture.

[3] H. Namari *et al.*, "Tube-type active scope camera with high mobility and practical functionality," in *IEEE/RSJ IROS 2012*, pp. 3679–3686.

[4] K. Ohno *et al.*, "Robotic control vehicle for measuring radiation in Fukushima Daiichi Nuclear Power Plant," in *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2011, pp. 38–43.

[5] K. Nagatani *et al.*, "Redesign of rescue mobile robot Quince," in *IEEE SSRR*, 2011, pp. 13–18.

[6] R. Voyles *et al.*, "Hexrotor UAV platform enabling dextrous interaction with structures – preliminary work," in *IEEE SSRR 2012*, pp. 1–7.

[7] V. Baiocchi *et al.*, "Development of a Software to Plan UAVs Stereoscopic Flight: An Application on Post Earthquake Scenario in LAquila Ciy," in *International Conference on Computational Science and Its Applications (ICCSA)*. Springer, 2013, pp. 150–165.

[8] S. Tadokoro *et al.*, "Application of active scope camera to forensic investigation of construction accident," in *IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, 2009, pp. 47–50.

[9] M. Ishikura *et al.*, "Shape estimation of flexible cable," in *IEEE/RSJ IROS 2012*, pp. 2539–2546.

[10] B. Jincun *et al.*, "The design of the rescue robot long-distance control based on 3G and GPS," in *International Conference on Intelligent Human-Machine Systems and Cybernetics, 2009*, vol. 2, pp. 170–172.

[11] S. Sukkarieh *et al.*, "A high integrity IMU/GPS navigation loop for autonomous land vehicle applications," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 3, pp. 572–578, 1999.

[12] Y. Kim *et al.*, "Thin polysilicon gauge for strain measurement of structural elements," *IEEE Sensors Journal*, vol. 10, no. 8, pp. 1320–1327, 2010.

[13] R. R. Murphy, *Disaster Robotics*. MIT Press, 2014.

[14] Y. Bando *et al.*, "Posture estimation of hose-shaped robot using microphone array localization," in *IEEE/RSJ IROS 2013*, pp. 3446–3451.

[15] D. Rosenthal *et al.*, *Computational auditory scene analysis*. CRC press, 1998.

[16] Y. Sasaki *et al.*, "Spherical microphone array for spatial sound localization for a mobile robot," in *IEEE IROS 2012*, pp. 713–718.

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

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

[19] E. A. Wan *et al.*, "The unscented kalman filter for nonlinear estimation," in *The IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.

[20] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[21] Y. Suzuki *et al.*, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *The Journal of the Acoustical Society of America*, vol. 97, p. 1119, 1995.

[22] C. Knapp *et al.*, "The generalized correlation method for estimation of time delay," *IEEE Trans. on ASSP*, vol. 24, no. 4, pp. 320–327, 1976.

[23] C. Zhang *et al.*, "Why does PHAT work well in lownoise, reverberative environments?" in *IEEE ICASSP 2008*, pp. 2565–2568.

[24] 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, 2010.

[25] R. van der Merwe *et al.*, "The unscented particle filter," in *Neural Information Processing Systems (NIPS)*, 2000, pp. 584–590.