IEICE TRANS. INF. & SYST., VOL.E94–D, NO.3 MARCH 2011

PAPER

# Position Identification by Actively Localizing Spacial Sound Beacons

Huakang LI[†], Jie HUANG[†a)], *and* Qunfei ZHAO[††], *Members*

**SUMMARY** In this paper, we propose a method for robot self-position identification by active sound localization. This method can be used for autonomous security robots working in room environments. A system using an AIBO robot equipped with two microphones and a wireless network is constructed and used for position identification experiments. Differences in arrival time to the robot's microphones are used as localization cues. To overcome the ambiguity of front-back confusion, a three-head-position measurement method is proposed. The position of robot can be identified by the intersection of circles restricted using the azimuth differences among different sound beacon pairs. By localizing three or four loudspeakers as sound beacons positioned at known locations, the robot can identify its position with an average error of 7 cm in a $2.5 \times 3.0\,m^2$ working space in the horizontal plane. We propose adjusting the arrival time differences (ATDs) to reduce the errors caused when the sound beacons are high mounted. A robot navigation experiment was conducted to demonstrate the effectiveness of the proposed position-identification system.

***key words:*** *sound based position identification, active robot position identification, adjusted pre-measurement ATDs, robot navigation by spatial sound localization*

## 1. Introduction

Self-position identification is very important for mobile robots. Many related studies have been conducted based on vision and image processing [1]–[3]. However, vision is not perfect. For example, in darkness or poor light, robots lose position and direction. When a home security robot is working at night in the absence of the owner, sound beacons are useful for robot self-position identification. Such sound-based position identification has advantages in poor lighting conditions and is robust against obstacles [4]–[6].

In this paper, we propose a new method for position identification using active sound localization for humanoid or pet robots like AIBO that are equipped with two microphones. The system requires three or more sound beacons located at different known positions. By actively localizing those sound beacons, the robot can identify its self-position. Here, the term "active" means that the robot can change head direction and control the sound beacons to emit sounds at any time.

The sound-based position identification system consists of two parts, sound source azimuth estimation and robot position identification.

Since the sound source can be actively controlled, we can choose a signal with one or several frequency components so that the sound energy will be concentrated and therefore robust against environment noise.

In developing sound-based robot sensing systems, it is beneficial to refer to the auditory systems of humans and other animals. The excellence of human and animal audition can serve as a good model for robot auditory systems. Inter-aural time difference (ITD), inter-aural intensity difference (IID), and head-related transfer function (HTRF) are important cues for sound localization in human audition [7], [8]. Of these, the ITD cue is easy to treat and has high localization accuracy. Similar to the ITD cue in human audition, the ATD between two microphones is used in this study for robot self-position identification. In general, the ATD between two microphones restricts a sound source to the surface of a rotated hyperbola. It can be approximated by the surface of a cone, known as the "cone of confusion" in psychoacoustic research [9], [10].

Two coordinated robot are used in the system. The sensing part is a pet robot, the AIBO robot provided by Sony, Inc. The AIBO robot has microphones located on each side of its face. With these microphones, the robot can obtain ATDs for sound signals. To avoid the cone of confusion, the azimuth of the sound source is determined by two or three trials, changing the direction of the robot head. We refer to this method as a two- or three-head-position method. By localizing three or four loudspeakers as sound beacons positioned at known locations, the robot can identify its position by the intersections restricted by the time differences among different microphone pairs. The motion part is a wheel-based mobile robot, LABO-3. By setting the AIBO on the LABO-3 at a height of 50 cm, the AIBO can inform the LABO-3 system of its current position to correct the position errors caused by the motion system.

In case there are difficulties to locate sound beacons at the same height as the robot microphones because of the positions of room furniture and other obstacles, we can mount the sound beacons at higher locations. The adjusted ATDs method is proposed to reduce the error caused by the elevation. Experiment results show that two- or three-head-position method can localize robot position in high accuracy even for the high mounted beacons.

Manuscript received July 8, 2010.
Manuscript revised October 29, 2010.

[†]The authors are with the School of Computer Science and Engineering University of Aizu, Aizuwakamatsu-shi, 965–8580 Japan.
[††]The author is with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, 200240, China.
a) E-mail: j-huang@u-aizu.ac.jp

DOI: 10.1587/transinf.E94.D.632

## 2. System Structure

The whole system is structured with the AIBO robot for sound recording, the LABO-3 robot for motion, and the PC client for calculation. The AIBO robot (Model ERS-210A/220A, Sony, Inc.) contains a 384 MHz processor and supports a wireless LAN. The microphones on the two sides of its head can receive stereo sound at a 16 kHz sampling rate. The head can turn left and right within 92.6 degrees in the horizontal axis, lean back within 46 degrees, and nod within 85 degrees in the vertical axis. As the moving speed of AIBO robot is relatively slow and is not accurate enough because of its leg-based moving mechanism. We use LABO-3, which is wheel-based and actuated by two motors with an output of 90 W and a maximum speed of 1.8 m/s, as the mobile subsystem.

In our system, the AIBO and LABO are connected with a PC by a wireless networks. Figure 1 depicts the processing flow between the PC server, and the AIBO and LABO-3 robots. For each measurement, the PC server sends a pure tone or a multi-frequency sound signal with 1 s duration time, to one of the selected sound beacons. Sound data is then recorded by the AIBO robot and sent to the server by wireless networks. After sound period extraction, the ATDs are calculated with cross-correlation or phase difference methods. The ATD measurements are performed for three times to estimate direction of a sound beacon (Sect. 4.2). Using the sound beacon pair selection method (Sect. 5.3), the robot position is identified by the intersection point of restricting circles which are obtained from the azimuth difference between two different sound beacons. With the identified position data, the remote PC performs the tasks of path planning, obstacle avoidance and robot navigation (Fig. 1).

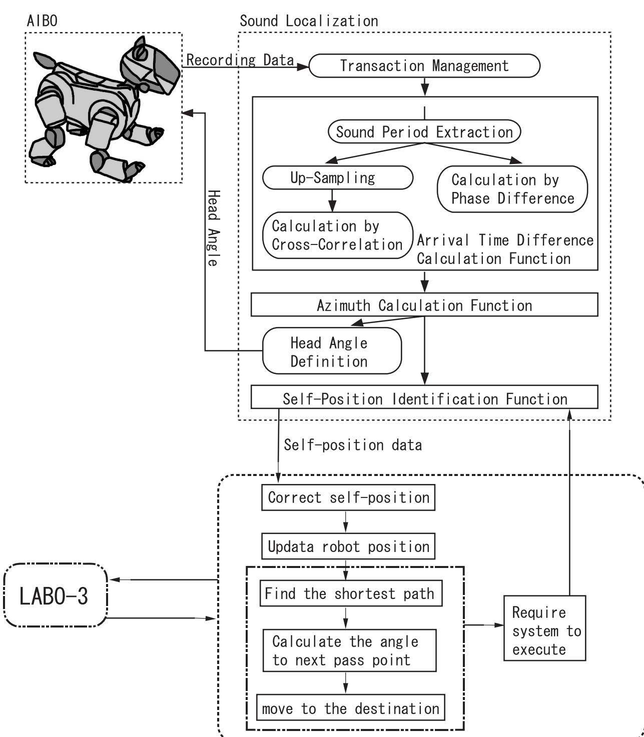


**Fig. 1** System structure and processing flow.

## 3. ATD Measurements

### 3.1 Pre-Measurement of ATDs

In order to estimate the sound source azimuth and identify the robot position, ATDs for different sound azimuths were pre-measured in an anechoic chamber ($5 \times 5 \times 5\,m^3$). The sound source was set at the same height as the two microphones and 1.5 m from the center of the robot head (Fig. 2). Empirically, when the sound distance is much bigger than the distance between the two microphones, the arrival time difference basically depends on the azimuth of sound source. Therefore, we set the azimuth $\theta$ from 0 to 180 degrees by a 5 degree step. The ATDs were calculated from the measured impulse responses. Figure 3 plots the pre-measured ATDs. The graph is nearly a sine curve but is not symmetric with respect to the front and back because of the asymmetry of the robot head and microphone locations, which are slightly toward the front.

### 3.2 Sound Period Extraction

Figure 4 plots sample data of both right and left micro-

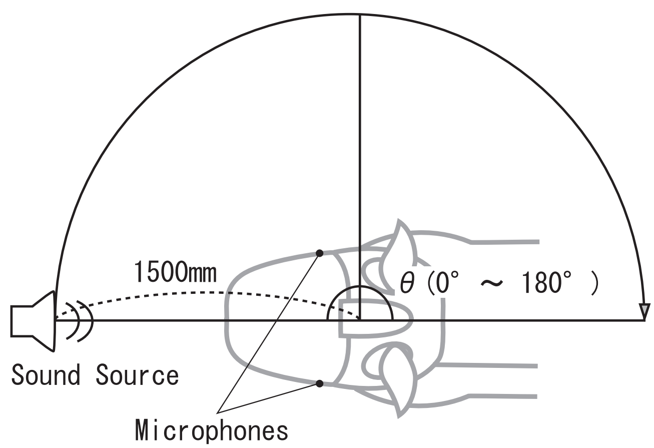


**Fig. 2** Setup for ATD pre-measurements.

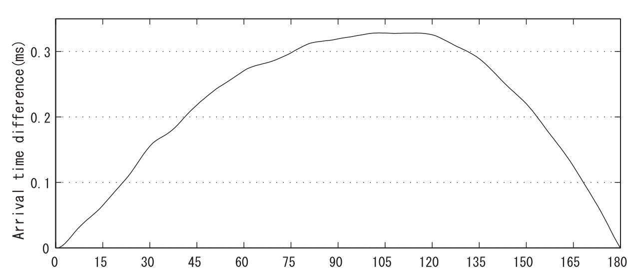


**Fig. 3** Pre-measured ATDs for azimuth 0 to 180 degree sound sources.

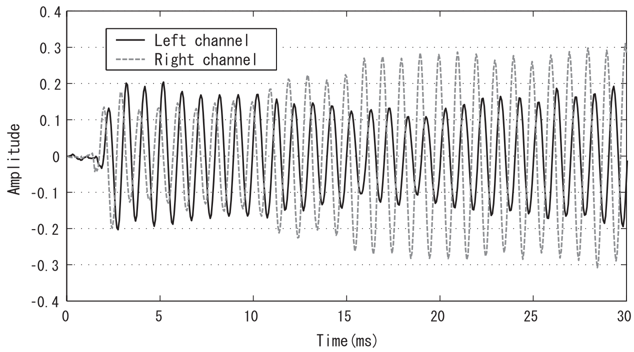


**Fig. 4** Left and right sound waves after a sound onset.

phones after a sound onset. The waveform disparities are not consistent; they are largely influenced by echoes and reverberations of the environment, except at their onset (a period of several milliseconds from the signal start point). In human auditory research, localizing sound sources with higher priority on their onset is known as the precedence effect [11], [12]. The echo estimation can be obtained by

$$e(t) = \begin{cases} 0 & 0 < t < \tau_0 \\ ke^{-(t-\tau_0)/\tau_d} & t \geq \tau_0 \end{cases} \tag{1}$$

where $\tau_0$ is the first reverberant time, and $k$ is the decay ratio of the ordinary room generated by learning algorithm. The sound onset was detected and extracted for ATD calculation.

### 3.3 ATD Calculation

The signals from two microphones are transformed by FFT, and the cross spectrum is calculated. To improve the resolution of the peak position, the FFT signals are zero-padded; thus, the sampling rate is increased by a factor of 16. The cross-correlation (CC) is then obtained to transform the signals back to the time domain by IFFT. The ATD is obtained from the peak position of the CC function.

Experiments of ATD calculation by phase differences were also conducted. The time delay $\Delta t$ can be obtained using the phase difference $\Delta\phi$:

$$\Delta t = \frac{\Delta\phi}{2\pi f} \ (-\pi \leq \Delta\phi \leq \pi) \tag{2}$$

where $f$ is the frequency of sound source. In this method, when there is more than one frequency, the common time difference for all frequency components is calculated. This method can also be used to avoid phase warping if there are high-frequency signals, such as the characteristic delay in animal audition [13].

Experiments of azimuth estimation were conducted in an ordinary room using sound stimuli of 1000 Hz pure tone and a mixture of 500, 1000 and 1200 Hz tones. The sound azimuths were calculated by both CC and phase difference (PD) methods (ignoring the front-back confusion discussed in the next section). For the PD method, the signals were filtered by a bandpass filter with the center frequency same as the sound signal to reduce environment noise. The results in Table 1 indicate that the average azimuth calculation errors were 3.2 to 3.3 degrees for both methods using the pure tone.

**Table 1** Average azimuth estimation errors (degrees) for all sound directions by cross-correlation (CC) and phase difference (PD) methods.

| | Pure Tone | Mixed sounds |
|---|---|---|
| CC | 3.23 | 3.29 |
| PD | 3.25 | 3.65 |

## 4. Azimuth Estimation

### 4.1 Two-Head-Position Method

As indicated by the pre-measured ATD graph, the sound source azimuth cannot be identified with just one measurement because of front-back confusion. Therefore, the robot changes the direction of its head and records sound two or more times. The first approach is to change the head 90 degrees to the right for a second measurement. The sound direction can then be judged by the intersection of two curves restricted by two ATDs.

Since azimuth sensitivity depends on the slope of the ATD curve, the accuracy of azimuth calculation is higher in high-slope areas (front and back) than in low-slope areas (left and right).

Azimuth estimation experiments were conducted in an anechoic chamber in order to analyze the azimuth estimation errors. The sound source direction was set from 0 to 360 degrees in 5-degree steps, and the average of azimuth errors was calculated for four estimates. In azimuth estimation, a smaller ATD yields a higher slope and accuracy; so the azimuth was calculated by the smaller ATD in the two measurements. Figure 5 indicates the average calculation errors for each azimuth.

The estimated azimuths have higher accuracy near 0, 90, 180, and 270 degrees because at such angles the robot head faces the sound source in at least one measurement. From this result we can expect the localization to be more precise if we turn the robot head to the sound source for the third measurement.

Another factor is the front-back difference. Table 2 presents the average errors for right-side (from 0 to 180 degrees) and left-side (from 180 to 360 degrees) sound sources. In the second measurement, the robot head turns to the right, and sound waves from the left side are influenced by the body of the robot. To overcome this drawback, we changed the method and turned the robot head 90 degrees to the sound source side in the second measurement. This improvement reduced the average azimuth calculation errors for both sides to 1.73 degrees.

### 4.2 Three-Head-Position Method

A three-head-position method was adopted to improve the

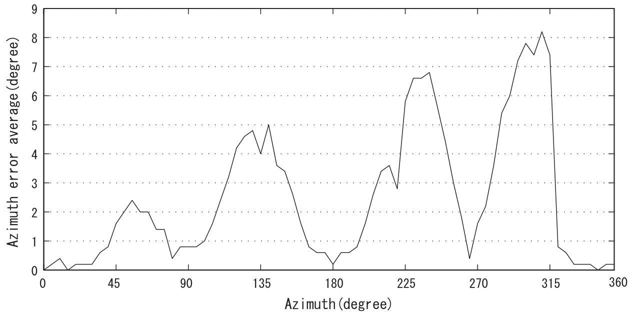


**Fig. 5** Azimuth estimation errors in an anechoic chamber using a two-head-position method (the second measurement was performed after turning the robot head 90 degrees to the right).

**Table 2** Average azimuth estimation errors (degrees) in an anechoic chamber of right- and left-side sound sources. The second measurement was performed after turning the head to the right or to the side of the sound source (s. s.).

| | Azimuth estimation error | |
|---|---|---|
| | turn right | turn to s. s. |
| Right side | 1.73 | 1.73 |
| Left side | 3.29 | 1.73 |

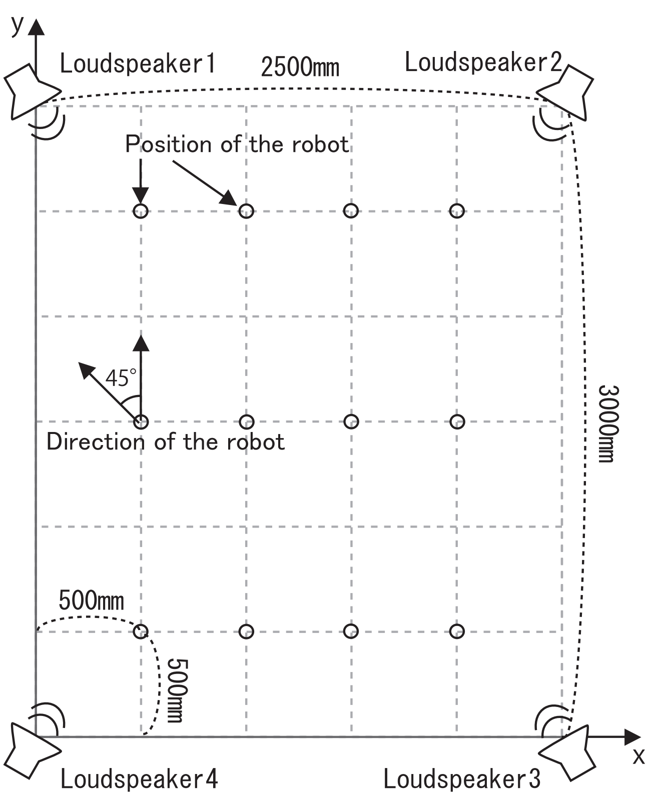


**Fig. 6** Experiment setup in a real environment (an ordinary room). The marks 'o' shows the measurement position. The direction of robot head was set to 0 and 45 degrees.

azimuth calculation error for sound sources at 45, 135, 225, and 315 degrees. With the two-head-position method, we obtained a relatively rough direction of the sound source. We then could turn the robot head toward the sound source

**Table 3** Errors of azimuth estimation in an ordinary room using different options for the third measurement.

| Options for backward | Azimuth estimation errors | | |
|---|---|---|---|
| | Overall | Front | Back |
| Opt. 1 | 2.31 | 1.46 | 3.17 |
| Opt. 2 | 2.83 | 1.46 | 4.21 |
| Opt. 3 | 1.81 | 1.46 | 2.17 |

to take a third measurement. Since the third measurement is performed in nearly the best condition, the robot is able to calculate the azimuth more precisely.

However, since the robot can turn its head only ±90 degrees, it cannot turn its head to sound sources behind it. There are three options with rearward sound sources:

1. turn the back of the head to the sound source.
2. avoid the third measurement and calculate the azimuth with only two measurements.
3. perform the third measurement if the azimuth of the sound source is between 115 and 155 degrees.

Experiments were conducted for these methods in the setup illustrated in Fig. 6. The robot was positioned in 12 locations in directions of 0 and 45 degrees (see Sect. 5 for details). The results presented in Table 3 indicate that the three-head-position method reduced the estimation errors, especially when the sound source was in front of the robot, and the best choice is option 3.

## 5. Active Robot Position Identification

### 5.1 Experimental Environment

The experiments were conducted in an open space ($2.5 \times 3.0m^2$) in an ordinary room, as depicted in Fig. 6. The walls, floor and ceiling of the room were not acoustically treated. The reverberation time is 36 ms and the average value of S/N is about 6.05 dB. Four sound beacons were arranged in the four corners at the same level as the microphones of the robot AIBO. The robot AIBO was positioned in 12 locations (circles in Fig. 6), in directions of 0 and 45 degrees. The robot received sound signals from the four sound beacons, and calculated the azimuth of each sound beacon by the methods described above. Then, from the estimated azimuths of sound beacons, the robot identified its self-position.

### 5.2 Identification Method

The azimuth information of one sound beacon is not enough to identify the self-position of the robot. Azimuths of two sound beacons can provide the azimuth difference ($\Delta\theta$) between the two sound beacons. With this information, the robot can restrict its position in a circle. If we can identify two circles by different sound beacon pairs, then the robot position can be identified as the intersection of the circles (Fig.7). Here, the intersection point at the position of the common sound beacon is eliminated.

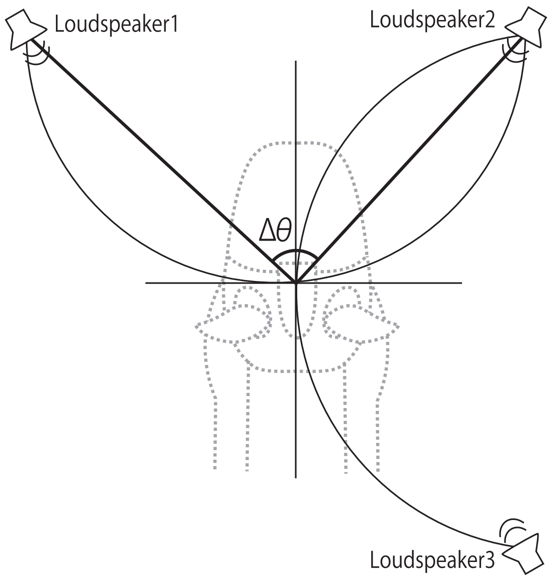


**Fig. 7** Circles restricted by ATDs and their intersection point.

In this identification process, if we localize the four sound beacons sequentially, we need to turn the robot head 12 times using the three-head-position method. However, we can do this in a parallel manner. First, roughly localize all the four sound beacons concurrently with the tow-head-position method. Then, turn to each sound beacon and perform the final measurement with the third head position measurement. Thus, the total number of robot head turning will reduce to six.

### 5.3 Sound Beacon Pair Selection Methods

Since we use a four-sound-beacon system, we can obtain a maximum of six different circles. In order to select microphones pairs for the best identification, three selection methods were compared in experiments.

1. *Four-sound-beacon average*: Use the average of intersections of six circles. There will be fifteen intersection points of different combinations.
2. *Obtuse angle priority*: Select the circles that form an obtuse angle $\Delta\theta$ from the robot position to the two sound beacons, because an obtuse angle is less error sensitive than an acute angle. The position of the robot is identified by averaging the selected circles.
3. *Front priority*: Select the circles obtained by front sound beacons, because the accuracy of azimuth estimation in front of the robot is higher than in back.

### 5.4 Identification Results

Table 4 shows position azimuth estimation errors by different sound beacon selection methods. From the results, the obtuse angle priority method has the highest accuracy than other methods.

The results of the active robot position identification,

**Table 4** Position identification errors (cm) by three microphone pair selection methods.

| sound beacon pair selection | Average errors |
|---|---|
| Average | 8.6 |
| Obtuse priority | 7.3 |
| Front priority | 8.8 |

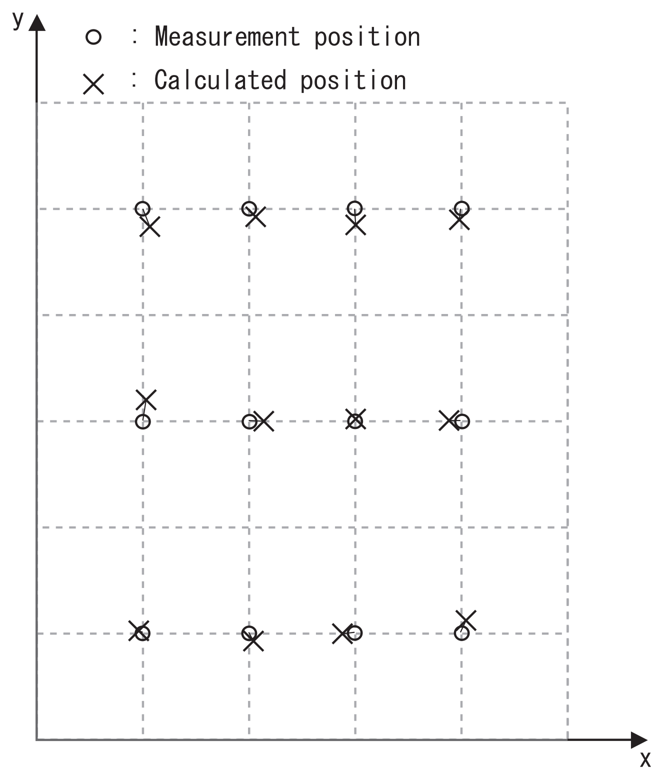


**Fig. 8** Results of position identification experiments. The identification results are denoted by X.

by the obtuse angle priority method is shown in Fig. 8. The error was about 7 cm in average.

## 6. Improvement of Position Identification in 3D Space Orientation

For the use of robots in home environments, sometimes it is difficult to locate sound beacons at the same height of the robot microphones because of the positions of room furniture and other obstacles. In these cases, we can mount the sound beacons higher than robot microphones with a height of $h$. Compared to the horizontal sound beacons, the ATD of a high mounted beacon will be reduced by a factor of $\cos\alpha$, where the elevation $\alpha$ depends on the horizontal distance $d$ from the robot to the sound beacon ($\alpha = \arctan\frac{h}{d}$).

### 6.1 Adjusted Pre-Measured ATDs

ATDs were pre-measured with the sound beacons placed at the same level as the robot's microphones in the horizontal plane. For accurate localization of the sound source by ATDs at a certain height, adjusted pre-measured ATDs were

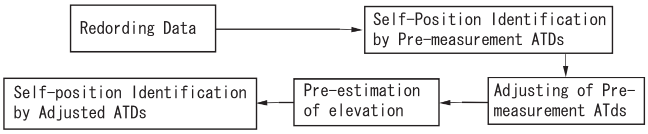


**Fig. 9** Position identification processing by adjusted pre-measurement ATDs.

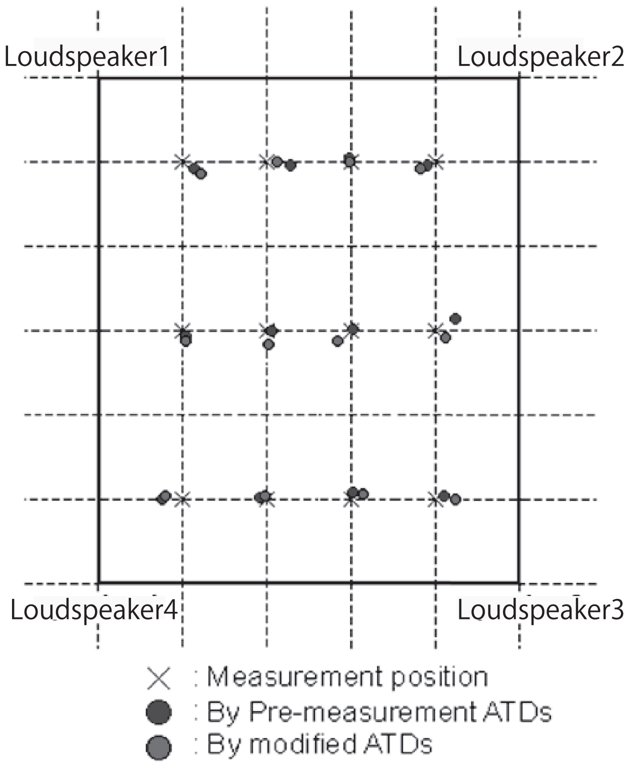


**Fig. 10** Results of position identification experiments by adjusted pre-measurement for the spacial environment.

proposed (as shown in Fig. 9):

$$\Delta t^{'} = \Delta t \cos \alpha \tag{3}$$

where $\Delta t^{'}$ is the adjusted pre-measured ATDs, $\Delta t$ is the pre-measured ATDs, and $\alpha$ is the angle of elevation.

### 6.2 Experiments and Results

The experiments were carried out in open space the same environment as mentioned in Sec.5.1, with the sound beacons placed 0.5 m above the horizontal plane of the robot's head. The robot's position identified by pre-measurement ATDs and adjusted ATDs are indicated in Fig. 10.

Table 5 presents the average difference between the actual position where the robot was placed and the position identified by sound localization with pre-measurement ATDs and adjusted ATDs. Even if the improvement of the adjusted was not significant, the adjustment approach confirmed that the head-related transfer function can improve the accuracy of identification.

**Table 5** Average errors (cm) of position identification with pre-measurement ATDs and adjusted pre-measurement ATDs.

| Method | Average errors |
|---|---|
| Pre-measurement ATDs | 8.58 |
| Adjusted ATDs | 8.11 |

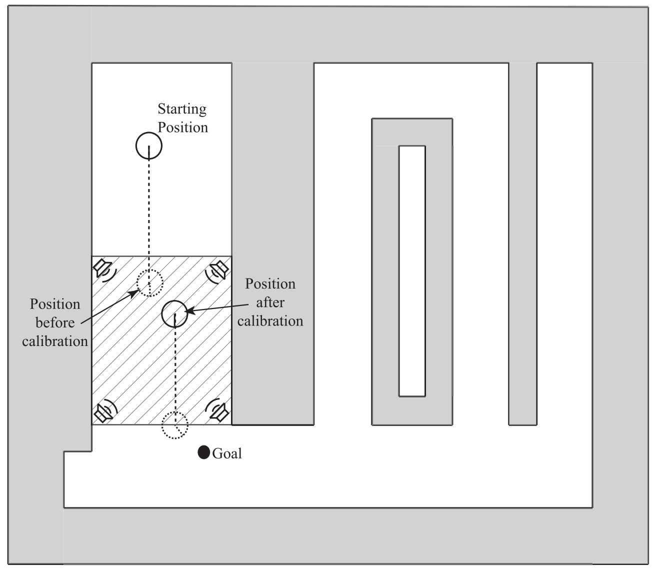


**Fig. 11** Robot navigation with sound-based position identification and calibration. The Gray areas indicate obstacles and other non-accessible fields. The unmarked (white) areas are the free areas for robot motion. The slash area is the sound-based position calibration area. The dotted lines represent the robot motion before and after position calibration.

**Table 6** Average distance errors to the destination with and without position calibration (five times average).

| | |
|---|---|
| without position calibration | 120 |
| with position calibration | 25 |

## 7. Robot Navigation and Position Identification

This experiment tested the robot navigation system with position identification in a room environment with a size of $11.5 \times 10\,m^2$ (Fig. 11). The reverberation time of the room was 475 ms and the average S/N was about 4.98 dB.

In this experiment, the robot was sitting in the initial position and moving to the goal. Because of the error of motion mechanism, there was an error between the real robot position and the robot position in the environment map. When the robot entered the position calibration area (the slashed area in Fig. 11), the robot performed position identification based on sound localization to calibrate its current position. After the calibration, the map position was updated. Then, the robot would find a new path from the updated position and finally approached to the goal.

Table 6 indicates the motion errors with and without position calibration by sound localization. The distance error to the destination was significantly corrected by the sound-based position calibration.

## 8. Conclusions

This paper proposed a sound-based robot position identification system based on active sound localization. In the horizontal plane, we analyzed the azimuth estimation errors produced by different methods of time difference calculation and sound source azimuth estimation. The best accuracy was improved to 1.8 degrees by a three-head-position measurement method. This improvement finally improved the performance of robot self-position identification to within 7 cm in a space of $2.5 \times 3.0\, m^2$. With the pre-adjusted ATDs, the system can locate its position to within 8 cm even in a spatial setting of sound beacons at a certain height. A robot navigation experiment was also conducted to demonstrate the effectiveness of the position-identification system.

The proposed sound-based robot position identification will be useful in various applications, including autonomous mobile or security guard robots.

### References

[1] F. Cuesta and A. Ollero, Intelligent mobile robot navigation, Springer Verlag, 2005.
[2] C. Zhao, Y. Ohtake, and J. Huang, "Robot position cal-ibration using colored rectangle signboards," J. Three Dimensional Images, vol.17, no.1, pp.166–169, 2003.
[3] J. Huang, C. Zhao, Y. Ohtake, H. Li, and Q. Zhao, "Robot position identification using specially designed landmarks," Instrumentation and Measurement Tech-nology Conference, IMTC 2006, Proc. IEEE, pp.2091–2094, 2006.
[4] J. Huang, N. Ohnishi, and N. Sugie, "Building ears for robots: sound localization and separation," Artificial Life and Robotics, vol.1, no.4, pp.157–163, 1997.
[5] J. Huang, N. Ohnishi, and N. Sugie, "Sound localization in reverberant environment based on the model ofthe precedence effect," IEEE Trans. Instrum. Meas., vol.46, no.4, pp.842–846, 1997.
[6] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, and N. Sugie, "A model-based sound local-ization system and its application to robot navigation," Robotics and Autonomous Systems, vol.27, no.4, pp.199–209, 1999.
[7] J. Blauert and J. Allen, Spatial hearing: the psychophysics of human sound localization, The MIT Press, 1997.
[8] D. Begault, 3-D sound for virtual reality and multimedia, Citeseer, 1994.
[9] S. Oldfield and S. Parker, "Acuity of sound localisation: a topography of auditory space. I. Normal hearing condi-tions," Perception, vol.13, no.5, pp.581–600, 1984.
[10] V. Algazi, R. Duda, R. Morrison, and D. Thompson, "Structural composition and decomposition of HRTFs," Proc. IEEE WASPAA01., pp.103–106, Citeseer, 2001.
[11] J. Huang, N. Ohnishi, and N. Sugie, "A biomimetic system for localization and separation of multiplesound sources," 1994 IEEE Instrumentation and Measurement Technology Conference, IMTC/94, Conference Proc. 10th Anniversary, Advanced Technologies in I & M, pp.967–970, 1994.
[12] J. Huang, N. Ohnishi, X. Guo, and N. Sugie, "Echo avoid-ance in a computational model of the precedence effect," Speech Commun., vol.27, no.3-4, pp.223–233, 1999.
[13] T. Takahashi and M. Konishi, "Selectivity for interaural time difference in the owl' s midbrain," J. Neuroscience, vol.6, no.12, p.3413, 1986.

**Huakang Li** was born in Suzhou, China in 1982, received a B.Eng. degree in School of Electric Information and Electrical Engineering from Shanghai Jiaotong University in 2005, the M.S. degree in computer and information system from the University of Aizu in 2007. He is currently prusuing the Ph.D. degree in the Ph.D. computer and information system at the University of Aizu, Japan.

**Jie Huang** was born in Wenzhou, China. He received B.Eng. degree from Nagoya Institute of Technology, Japan, in 1985. He received M. Eng nad D.Eng. degrees from Nagoya University in 1987 and 1991. From 1992 to 1993 he was a system engineer at Phenix Data Corp. He worked as a frontier research scientist at the Bio-Mimetic Control Research Center under the Institute of Physical and Chemical Research (RIKEN), Japan, from 1993 to 1998. He was also a visiting researcher at GMD, German National Research Center for Information Technology engaged on Cognitive Robotics Research from Dec. 1997 to Jan. 1998. He was an assistant professor at the University of Aizu, in the Multimedia Devices Lab, School of Computer Science and Engineering, from 1998 to 2002. Since 2002, he has been an associate professor at the University of Aizu, in the Human Interface Lab, School of Computer Science and Engineering.

**Qunfei Zhao** received B.S. degree from Xi'an Jiaotong University at Xi'an, China in 1982. He received the M.S. and D.S. degrees from Tokyo Institute of Technology at Tokyo, Japan in 1985 and 1988. From Apr. 1988 to May. 1989 he was a researcher at Mitsubishi Research Institute. Inc.. He worked as a engineer at Research Institute of Economy, Trade and Industry from Jun. 1989 to Mar. 1991. He was also an associate professor at North Jiaotong University in City Traffic Lab, System Analysis Institute from Mar. 1991 to Mar. 1994. He also worked as a chief of department of aviation institution in Kawada Industries, Inc. from Apr. 1994 to May. 2001. He worked as a professor and the associate head of Robotic institute at Shanghai Jiaotong University from Jun. 2001 to Feb. 2004. Since Mar. 2004, he has been a professor and tutor of Ph.D. supervisor in the Institute of Image Processing and Pattern Identification. From Feb. 2008, he has been an visiting professor at Waseda University.