We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Sound Source Localization Method Using Region Selection

Yong-Eun Kim[1], Dong-Hyun Su[2], Chang-Ha Jeon[2], Jae-Kyung Lee[2],
Kyung-Ju Cho[3] and Jin-Gyun Chung[2]
[1]Korea Automotive Technology Institute in Chonan,
[2]Chonbuk National University in Jeonju,
[3]Korea Association Aids to Navigation in Seoul,
Korea

1. Introduction

There are many applications that would be aided by the determination of the physical position and orientation of users. Some of the applications include service robots, video conference, intelligent living environments, security systems and speech separation for hands-free communication devices (Coen, 1998; Wax & Kailath, 1983; Mungamuru & Aarabi, 2004; Sasaki et al., 2006; Lv & Zhang 2008). As an example, without the information on the spatial location of users in a given environment, it would not be possible for a service robot to react naturally to the needs of the user.

To localize a user, sound source localization techniques are widely used (Nakadai et al., 2000; Brandstein & Ward, 2001; Cheng & Wakefield, 2001; Sasaki et al., 2006). Sound localization is the process of determining the spatial location of a sound source based on multiple observations of the received sound signals. Current sound localization techniques are generally based upon the idea of computing the time difference of arrival (TDOA) information with microphone arrays (Knnapp & Cater, 1976; Brandstein & Silverman, 1997).

An efficient method to obtain TDOA information between two signals is to compute the cross-correlation of the two signals. The computed correlation values give the point at which the two signals from separate microphones are at their maximum correlation. When only two isotropic (i.e., not directional as in the mammalian ear) microphones are used, the system experiences front-back confusion effect: the system has difficulty in determining whether the sound is originating from in front of or behind the system. A simple and efficient method to overcome this problem is to incorporate more microphones (Huang et al., 1999).

Various weighting functions or pre-filters such as Roth, SCOT, PHAT, Eckart filter and HT can be used to increase the performance of time difference estimation (Knnapp & Cater, 1976). However, the performance improvement is achieved with the penalty of large power consumption and hardware overhead, which may not be suitable for the implementation of portable systems such as service robots.

In this chapter, we propose an efficient sound source localization method under the assumption that three isotropic microphones are used to avoid the front-back confusion

effect. By the proposed approach, the region from 0° to 180° is divided into three regions and only one of the three regions is selected for the sound source localization. Thus considerable amount of computation time and hardware cost can be reduced. In addition, the estimation accuracy is improved due to the proper choice of the selected region.

2. Sound localization using TDOA

If a signal emanated from a remote sound source is monitored at two spatially separated sensors in the presence of noise, the two monitored signals can be mathematically modeled as

$$
\begin{aligned}
x_1(t) &= s_1(t) + n_1(t), \\
x_2(t) &= \alpha s_1(t - D) + n_2(t),
\end{aligned}
\tag{1}
$$

where α and D denote the relative attenuation and the time delay of $x_2(t)$ with respect to $x_1(t)$, respectively. It is assumed that signal $s_1(t)$ and noise $n_i(t)$ are uncorrelated and jointly stationary random processes. A common method to determine the time delay D is to compute the cross correlation

$$
R_{x_1 x_2}(\tau) = E[x_1(t)x_2(t - \tau)] ,
\tag{2}
$$

where E denotes expectation operator. The time argument at which $R_{x_1 x_2}(\tau)$ achieves a maximum is the desired delay estimate.
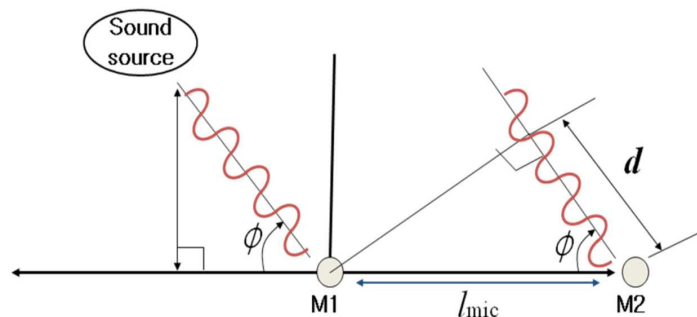

Fig. 1. Sound source localization using two microphones

Fig. 1 shows the sound localization test environments using two microphones. We assume that the sound waves arrive in parallel to each microphone as shown in Fig. 1. Then, the time delay D can be expressed as

$$
D = \frac{d}{v_{sound}} = \frac{l_{mic} \cos \phi}{v_{sound}} ,
\tag{4}
$$

where v_{sound} denotes the sound velocity of 343m/s. Thus, the angle of the sound source is computed as

$$
\phi = \cos^{-1} \frac{D \, v_{sound}}{l_{mic}} = \cos^{-1} \frac{d}{l_{mic}} .
\tag{5}
$$

If the sound wave is sampled at the rate of f_s, and the sampled signal is delayed by n_d samples, the distance d can be computed as

$$d = \frac{v_{sound} n_d}{f_s} . \tag{6}$$

In Fig. 1, since d is a side of a right-angled triangle, we have

$$d < l_{mic} . \tag{7}$$

Thus, when $d = l_{mic}$ in (6), the number of maximum delayed samples $n_{d,\max}$ is obtained as

$$n_{d,\max} = \frac{f_s l_{mic}}{v_{sound}} . \tag{8}$$

3. Proposed sound source localization method

3.1 Region selection for sound localization

The desired angle in (5) is obtained using the inverse cosine function. Fig. 2 shows the inverse cosine graph as a function of d. Since the inverse cosine function is nonlinear, Δd (estimation error in d) has different effect on the estimated angle depending on the sound source location. Fig. 3 shows the estimation error (in degree) of sound source location as a function of Δd. As can be seen from Fig. 3, Δd has smaller effect for the sources located from 60° to 120°. As an example, when the source is located at 90° with the estimation error $\Delta d = 0.01$, the mapped angle is 89.427°. However, if the source is located at 0° with the estimation error $\Delta d = 0.01$, the mapped angle is 8.11°. Thus, for the same estimation error Δd, the effect for the source located at 0° is 14 times larger than that of the source at 90°. To efficiently implement the inverse cosine function, we consider the region from 60° to 120° as approximately linear as shown in Fig. 2.
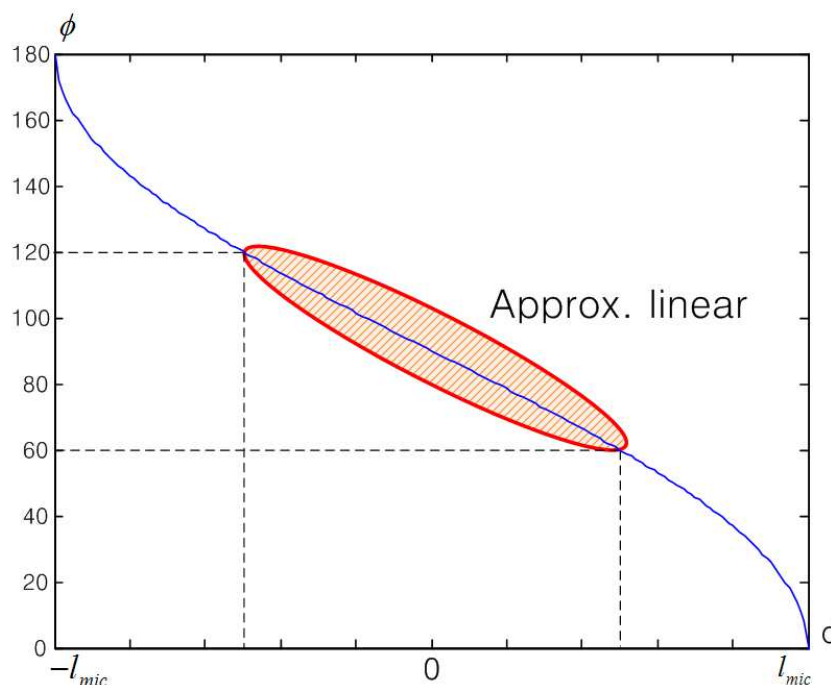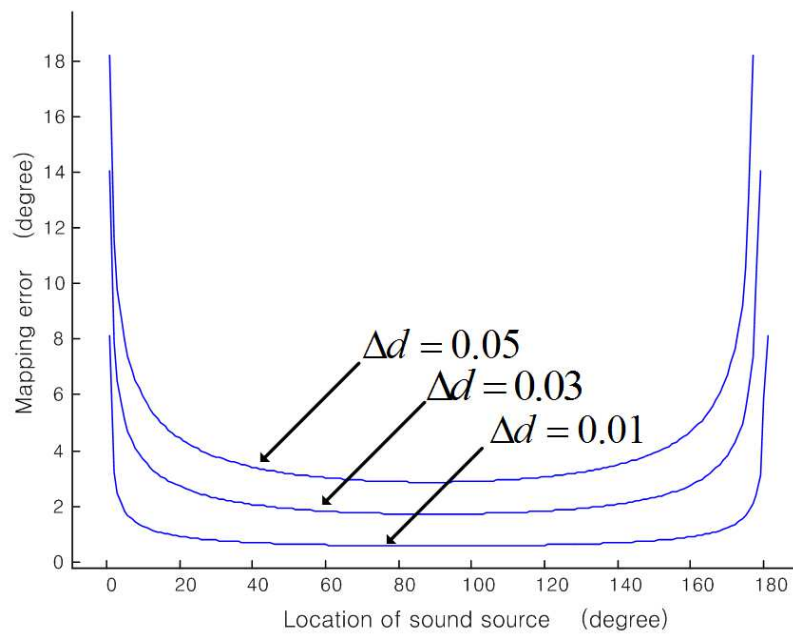

Fig. 2. Inverse cosine graph as a function of d

Fig. 3. Estimation error of sound source location as a function of Δd

Fig. 4 shows the front-back confusion effect: the system has difficulty in determining whether the sound is originating from in front of (sound source A) or behind (sound source B) the system. A simple and efficient method to overcome this problem is to incorporate more microphones. In Fig. 5, three microphones are used to avoid the front-back confusion effect, where L, R and B mean the microphones located at the left, right and back sides, respectively. In this chapter, to apply the cross-correlation operation in (2), for each arrow between the microphones in Fig. 5, the signal received at the tail part and the head part are designated as $x_1(t)$ and $x_2(t)$, respectively.

In conventional approaches, correlation functions are calculated between each microphone pair and mapped to angles as shown in Fig. 6-(a), (b) and (c). Notice that, due to the front-back confusion effect, each microphone pair provides two equivalent maximum values. Fig. 6-(d) is obtained by adding the three curves. In Fig. 6-(d), the angle corresponding to the maximum magnitude is the desired sound source location.
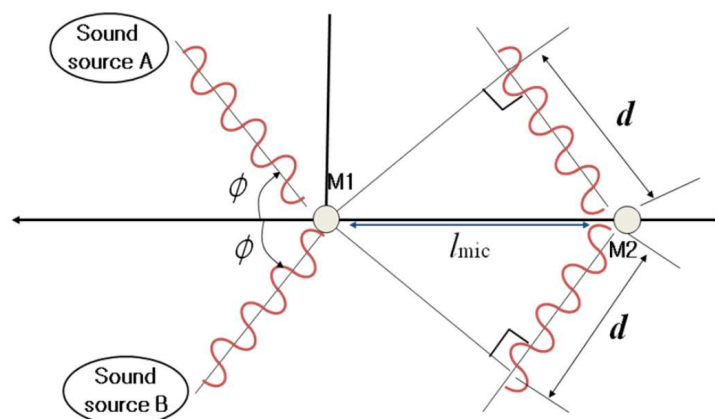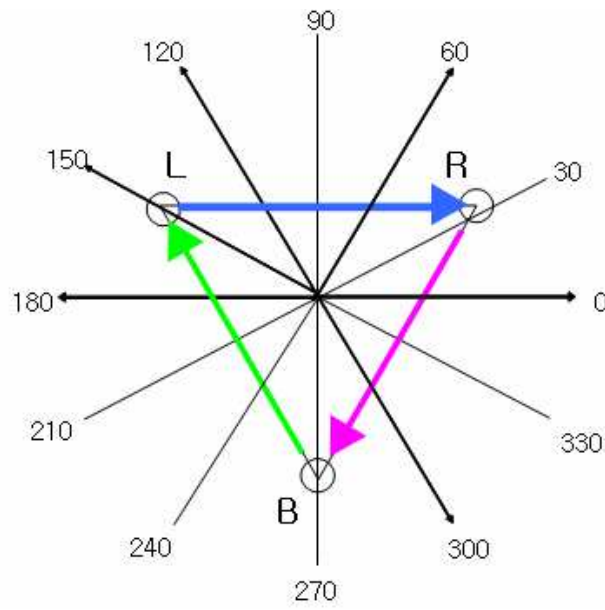

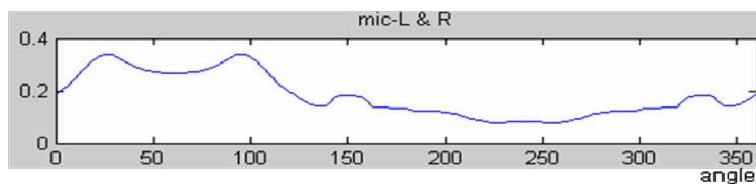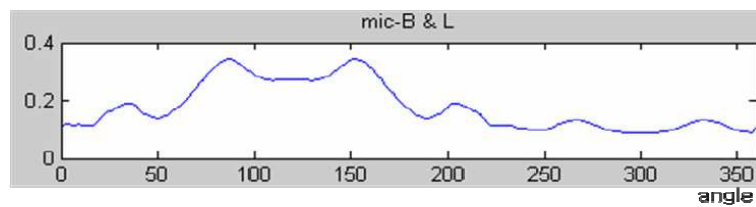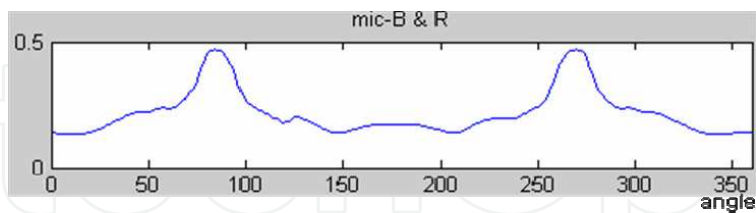Fig. 4. Front-back confusion effect

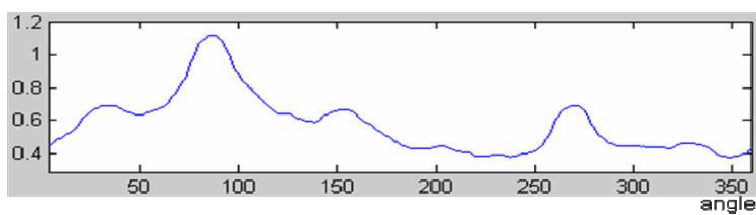Fig. 5. Sound source localization using three microphones

(a)

(b)

(c)

(d)

Fig. 6. Angles obtained from microphone pairs: (a) L-R, (b) B-L, (c) R-B, and (d) (L-R)+ (B-L)+(R-B)

Source location(angle)	Proper microphone pair
60°∼120°, 240°∼300°	R-L
120°∼180°, 300°∼360°	B-R
180°∼240°, 0°∼60°	L-B

Table 1. Selection of proper microphone pair for six different source locations.

Due to the nonlinear characteristic of the inverse cosine function, the accuracy of each estimation result is different depending on the source location. Notice that in Fig. 5, wherever the source is located, exactly one microphone pair has the sound source within its approximately linear region (60°~120° or 240°~300° for the microphone pair). As an example, if a sound source is located at 30° in Fig. 5, the location is within the approximately linear region for L-B pair. Table 1 summarizes the choice of proper microphone pairs for six different source locations.

The proper selection of microphone pairs can be achieved by comparing the time index τ_{max} values (or, the number of shifted samples) in (2) at which the maximum correlation values are obtained. Fig. 7 shows the comparison of the correlation values obtained from three microphone pairs when the source is located at 90°. For the smallest estimation error, we select the microphone pair whose τ_{max} value is closest to 0. Notice that the correlation curve in the center (by the microphone pair R-L) has the τ_{max} value which is closest to 0.

In fact, for the smallest estimation error, we just need to select the correlation curve in the center. As an example, assume that a sound source is located at 90° in Fig. 5. Then, for the microphone pair R-L, the two signals arrived at the microphones R and L have little difference in their arrival times since the distances from the source to each microphone are almost the same. Thus, the cross correlation has its maximum around $\tau = 0$. However, for L-B pair, the microphone L is closer to the source than the microphone B. Since the received signals at microphones B and L are designated as $x_1(t)$ and $x_2(t)$, respectively, the cross
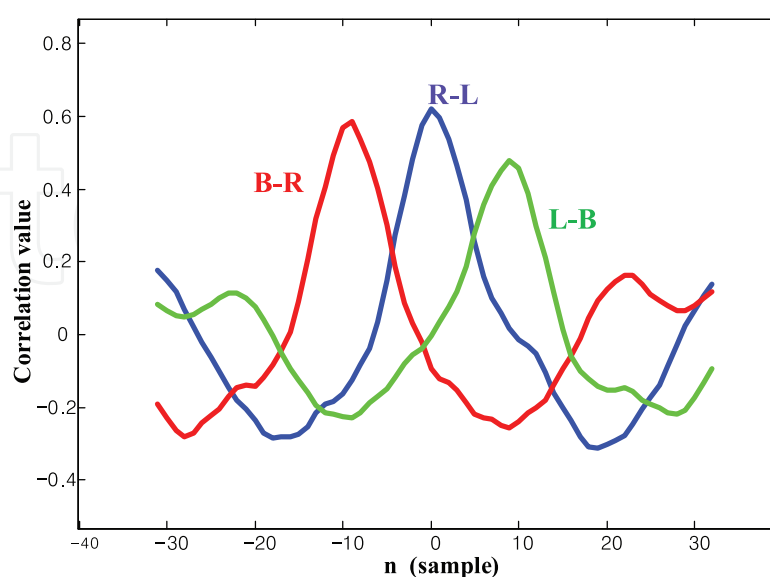

Fig. 7. Comparison of the correlation values obtained from three microphone pairs for the source located at 90°

correlation in (2) gets its maximum when $x_2(t)$ is shifted to the right ($\tau > 0$). The opposite is true for the microphone pair B-R as can be seen from Fig. 7.

Table 2 shows that proper microphone pairs can be simply selected by comparing maximum correlation positions (or, $\tau_{\max}$ values from each microphone pair).

Maximum correlation positions	Proper Mic.	Front / Back
$\tau_{\max}$ (BR)≤ $\tau_{\max}$ (RL) ≤ $\tau_{\max}$ (LB)	R-L	Front
$\tau_{\max}$ (BR)≤ $\tau_{\max}$ (LB) ≤ $\tau_{\max}$ (RL)	L-B	Front
$\tau_{\max}$ (RL)≤ $\tau_{\max}$ (BR) ≤ $\tau_{\max}$ (LB)	B-R	Front
$\tau_{\max}$ (LB)≤ $\tau_{\max}$ (RL) ≤ $\tau_{\max}$ (BR)	R-L	Back
$\tau_{\max}$ (RL)≤ $\tau_{\max}$ (LB) ≤ $\tau_{\max}$ (BR)	L-B	Back
$\tau_{\max}$ (LB)≤ $\tau_{\max}$ (BR) ≤ $\tau_{\max}$ (RL)	B-R	Back

Table 2. Selecetion of proper microphone pair

If the sampled signals of $x_1(t)$ and $x_2(t)$ are denoted by two vectors X_1 and X_2, the length of the cross-correlated signal R_{X1X2} is determined as

$$n(R_{X1X2}) = n(X_1) + n(X_2) - 1, \tag{9}$$

where $n(X)$ means the length of vector X. In other words, to obtain the cross-correlation result, vector shift and inner product operations need to be performed by $n(R_{X1X2})$ times.

It is interesting to notice that, once the distance between the microphones and the sampling rate are determined, the maximum time delay between two received signals is bounded by $n_{d,\max}$ in (8). Thus, instead of performing vector shift and inner product operations by $n(R_{X1X2})$ times as in the conventional approaches, it is sufficient to perform the operations by only $n_{d,\max}$ times. Specifically, we perform the correlation operation from $n = -n_{d,\max}/2$ to $n = n_{d,\max}/2$ (for sampled signals, $\tau = n/f_s$, integer n). In the simulation shown in Fig. 7, $n(X_1) = n(X_2) = 256$ and $n_{d,\max} = 64$. Thus, the number of operations for cross-correlation is reduced from 511 to 65 by the proposed method, which means the computation time for cross-correlation can be reduced by 87%.

3.2 Simplification of angle mapping using linear equation

Conventional angle mapping circuits require a look-up table for inverse cosine function. Also, an interpolation circuit is needed to obtain a better resolution with reduced look-up table. However, since the proposed region selection approach uses only the approximately linear part of the inverse cosine function, the use of look-up table and interpolation circuit can be avoided. Instead, the approximately linear region is approximated by the following equation:

$$y = ax + b, \tag{10}$$

where

$$a = \frac{-60}{(\cos\pi/3 - \cos2\pi/3) \times l_{mic}},$$
$$b = 120 + \frac{60\cos2\pi/3}{(\cos\pi/3 - \cos2\pi/3)}.$$

(11)

When the distance between the two microphones is given, the coefficients a and b in (10) can be pre-calculated. Thus, angle mapping can be performed using only one multiplication and one addition for a given value of d.

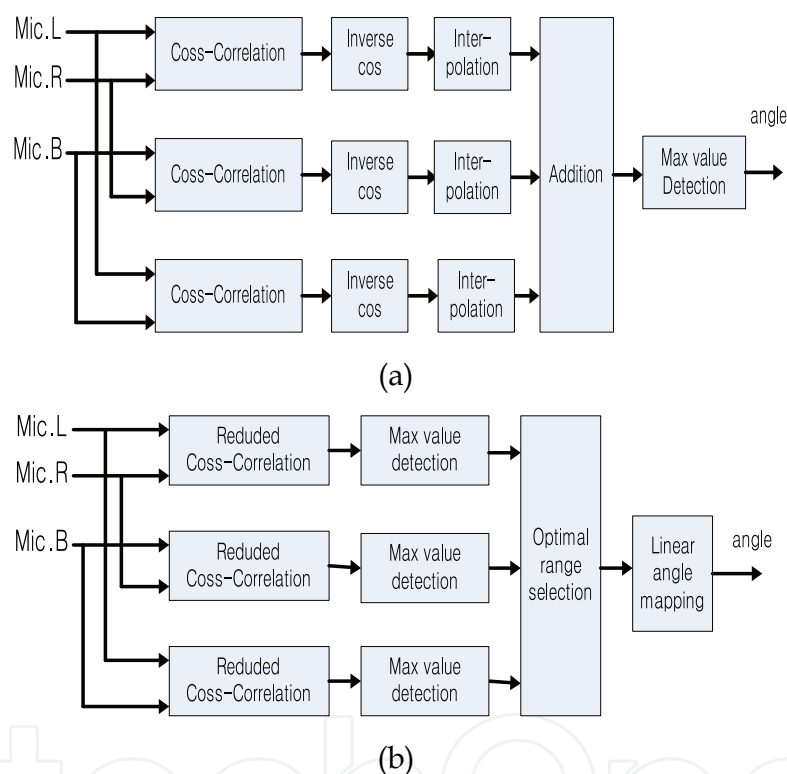Fig. 8 shows the block diagrams of the conventional sound source localization systems and the proposed system.

(a)

(b)

Fig. 8. Block diagrams of conventional and proposed methods: (a) conventional method, and (b) proposed method.

4. Simulation results

Fig. 9 shows the sound source localization system test environments. The distance between the microphones is 18.5cm. The sound signals received using three microphones are sampled at 16 KHz and the sampled signals are sent to the sound localization system implemented using Altera stratix II FPGA. Then, the estimation result is transmitted to a host PC through two FlexRay communication systems. The test results are shown in Table 3. Notice that the average error of the proposed method is only 31% of that of the conventional method. To further reduce the estimation error, we need to increase the sampling rate and the distance between the microphones.
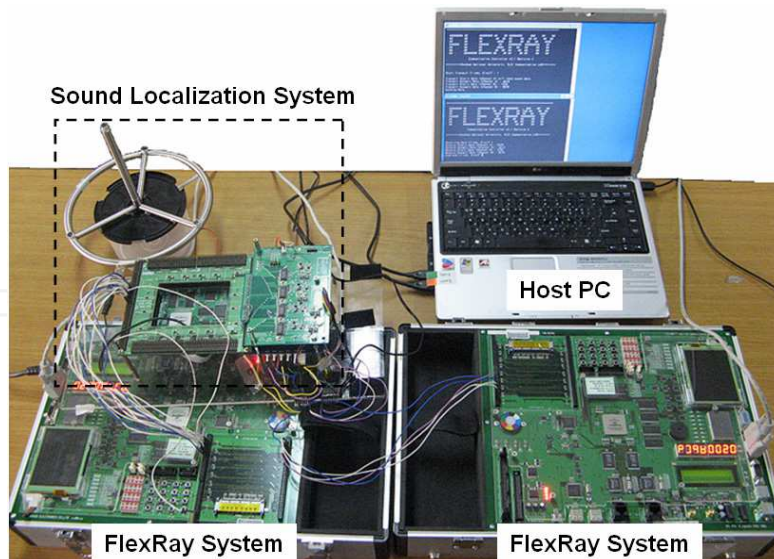
Fig. 9. Sound localization system test environments

Distance	0°	30°	60°	90°
1m	0°	27°	56°	88°
2m	0°	27°	59°	85°
3m	0°	27°	59°	88°
4m	2.5°	34°	57°	95°
5m	4.1°	37°	67°	82°
Maximum absolute error	4.1°	7°	7°	8°
average error	1.32°	4°	3.2°	4.4°

(a)

Distance	0°	30°	60°	90°
1m	0°	32.7°	60°	87.2°
2m	0°	32°	59°	85°
3m	0°	32.7°	60°	87.2°
4m	1	28°	62°	86°
5m	2	33°	61°	92°
Maximum absolute error	2°	3°	2°	4°
average error	0.6°	2.48°	0.8°	3.32°

(b)

Table 3. Simulation results: (a) conventional method, and (b) proposed method

5. Conclusion

Compared with conventional sound source localization methods, proposed method achieves more accurate estimation results with reduced hardware overhead due to the new region selection approach. By the proposed approach, the region from 0° to 180° is divided into three regions and only one of the three regions is selected such that the selected region corresponds to the linear part of the inverse cosine function. By the proposed approach, the

computation time for cross correlation is reduced by 87%, compared with the conventional approach. By simulations, it is shown that the estimation error by the proposed method is only 31% of that of the conventional approach.

The proposed sound source localization system can be applied to the implementation of portable service robot systems since the proposed system requires small area and low power consumption compared with conventional methods. The proposed method can be combined with generalized correlation method with some modifications.

6. Acknowledgment

7. References

Brandstein M. S. & Silverman H. (1997). A practical methodology for speech source localization with microphone arrays. *Comput. Speech Lang.*, Vo.11, No.2, pp. 91-126, ISSN 0885-2308

Brandstein M. & Ward D. B. (2001). Robust *Microphone Arrays: Signal Processing Techniques and Applications*, New York: Springer, ISBN 978-3540419532

Cheng I. & Wakefield G. H. (2001). Introduction to head-related transfer functions (HRTFs): representations of HRTFs in time, frequency, and space. *J. Audio Eng. Soc.*, Vol. 49, No.4, (April, 2001), pp. 231-248, ISSN 1549-4950

Coen M. (1998). Design principles for intelligent environments, *Proceedings of the 15th National Conference on Artificial Intelligence*, pp. 547-554

Huang J.; Supaongprapa T.; Terakura I.; Wang F.; Ohnishi N. & Sugie N. (1999) A model-based sound localization system and its application to robot navigation. *Robot. Auton. Syst.*, Vol.27, No.4, (June,1999), pp. 199-209, ISSN 0921-8890

Knnapp C. H. & Cater G. C. (1976). The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust. Speech Signal Process.*, Vol.24, No.4, (August 1976), pp.320-327, ISSN 0096-3518

Lv X. & Zhang M. (2008). Sound source localization based on robot hearing and vision, *Proceedings of ICCSIT 2008 International Conference of Computer Science and Information Technology*, pp. 942-946, ISBN 978-0-7695-3308-7, Singapore, August 29-September 2 2008

Mungamuru, B. & Aarabi, P. (2004). Enhanced sound localization. *IEEE Trans. Syst. Man Cybern. Part B- Cybern.*, Vol.34, No.3, (June, 2004), pp. 1526-1540, ISSN 1083-4419

Nakadai K.; Lourens T.; Okuno H. G. & Kitano H. (2000). Active audition for humanoid, *Proceedings of the 17th National Conference on Artificial Intelligence* and *12th Conference on Innovative Applications of Artificial Intelligence*, pp. 832-839

Sasaki Y.; Kagami S. & Mizoguchi H. (2006). Multiple sound source mapping for a mobile robot by self-motion triangulation, *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 380-385, ISBN 1-4244-0250-X, Beijing, China, October, 2006

Wax M. & Kailath T. (1983). Optimum localization of multiple sources by passive arrays. *IEEE Trans. Acoust. Speech Signal Process.*, Vol.31, No.6, (October,1983). pp. 1210-1217, ISSN 0096-3518
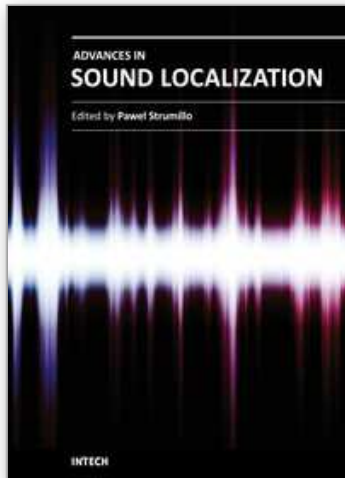
Advances in Sound Localization

Edited by Dr. Pawel Strumillo

ISBN 978-953-307-224-1

Hard cover, 590 pages

Publisher InTech

Published online 11, April, 2011

Published in print edition April, 2011

Sound source localization is an important research field that has attracted researchers' efforts from many technical and biomedical sciences. Sound source localization (SSL) is defined as the determination of the direction from a receiver, but also includes the distance from it. Because of the wave nature of sound propagation, phenomena such as refraction, diffraction, diffusion, reflection, reverberation and interference occur. The wide spectrum of sound frequencies that range from infrasounds through acoustic sounds to ultrasounds, also introduces difficulties, as different spectrum components have different penetration properties through the medium. Consequently, SSL is a complex computation problem and development of robust sound localization techniques calls for different approaches, including multisensor schemes, null-steering beamforming and time-difference arrival techniques. The book offers a rich source of valuable material on advances on SSL techniques and their applications that should appeal to researches representing diverse engineering and scientific disciplines.

INTECH
open science | open minds