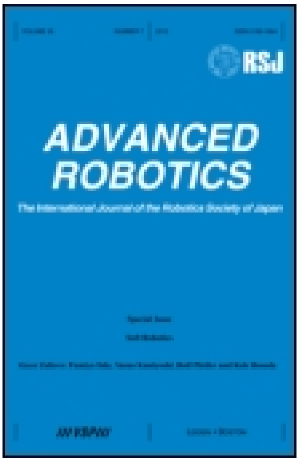


Advanced Robotics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tadr20

Gaussian mixture-sound field landmark model for robot localization applications

Li-Wei Wu [a], Wei-Han Liu [b], Chieh-Cheng Cheng [c] & Jwu-Sheng Hu [d]

[a] Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan
[b] Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan
[c] Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan
[d] Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan

Published online: 02 Apr 2012.

To cite this article: Li-Wei Wu , Wei-Han Liu , Chieh-Cheng Cheng & Jwu-Sheng Hu (2007) Gaussian mixture-sound field landmark model for robot localization applications, Advanced Robotics, 21:5-6, 619-643, DOI: 10.1163/156855307780108259

To link to this article: http://dx.doi.org/10.1163/156855307780108259

Advanced Robotics, Vol. 21, No. 5–6, pp. 619–643 (2007)

Also available online - www.brill.nl/ar

Full paper

Gaussian mixture-sound field landmark model for robot localization applications

LI-WEI WU *, WEI-HAN LIU, CHIEH-CHENG CHENG and JWU-SHENG HU

Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan

Received 9 February 2006; accepted 4 July 2006

Abstract—This work proposes a novel Gaussian mixture-sound field landmark model for localization applications, based on the principle that sound fields produced by sources at different locations can be distinguished in terms of their statistical patterns. The experimental results indicate that two microphones are sufficient to differentiate among the patterns. The proposed method is robust against environmental noise and performs accurately in a complex environment. Moreover, it cannot only detect the non-line-of-sight locations when the direct path between the microphones and the location is blocked, but also can distinguish the locations aligned with respect to the line connecting the microphones. However, using only two microphones, these scenarios are difficult to handle by traditional direction-of-arrival or beamforming methods in microphone array research. The experiments were conducted on a quadruped robot platform with an *e*Robot agent using embedded Ethernet technology. Because of its high accuracy and low-cost, this method is suitable for robot localization in real environments. The experimental results also show that the proposed method with only two microphones outperforms the conventional multiple signal classification method (MUSIC) technique with six microphones at various signal-to-noise ratios.

Keywords: Gausian mixture model; robot; sound field landmark; localization.

1. INTRODUCTION

Self-localization is one of the most important technologies for autonomous navigation in robotics [1, 2]. The sensor cost, system flexibility, shelter effect (non-line-of-sight) and precision are important issues in robot localization [3]. Localization technologies have been developed for indoor mobile robots [1, 2] based on various sensors, such as the inertial navigator sensor, ultrasonic sensor array, omnidirectional image sensor, radio frequency identification (RFID), wireless local area

[*] To whom correspondence should be addressed. Current address: New Technology Development Department, Compal Communications, Nanjing, Taiwan. E-mail: liwei.ece89g@nctu.edu.tw

network (WLAN) and infrared (IR). Inertial navigation sensors utilize gyroscopes, accelerate sensors and odometer recorders to calculate the robot's location. However, the inertial navigator has inherent error accumulation problems, and typically requires assistance from other sensors such as an ultrasonic sensor array, laser range scanner or IR to compensate for the system errors and to provide the initial location information [4–8]. For instance, Saito [9] proposed a teaching and playback navigation system which records data from an odometer and an ultrasonic range sensor. Ohya *et al.* [10] presented a navigation method using a camera and an ultrasonic sensor. However, the shelter effect occurs when the direct path from the robot to the sensor is blocked and could significantly influence the accuracy of these sensors. Methods based on WLAN [11, 12] and RFID [13] that can overcome shelter effect have been proposed. The WLAN-based methods [11, 12] require more than three access point (AP) stations in the operation zone, and their accuracies are currently limited to 1.5 and 2 m. These results are insufficiently accurate for general indoor applications. In Ref. [13], a RFID-based technology was demonstrated to achieve simultaneous localization and mapping using massive deployed RFID tags in space. However, tag cost, receiving range, antenna size, convenience of deployment and accuracy still need further investigations [14]. Lately, many researchers have employed omnidirectional image sensors, which can acquire a 360° view around a robot, for their navigation systems [15]. For example, Matsumoto [16] proposed a teaching and playback navigation method using a memorized omnidirectional view sequence, and attained satisfactory experimental results in an indoor environment. In Matsumoto's method, the robot only used its views to match the memorized views and decide its motion. However, the omnidirectional image sensor is more expensive than the sensors mentioned above. Additionally, vision algorithms always need more memory and computational effort than other methods mentioned above.

The characteristics of sound have already been used for localization or navigation by animals, such as bats and dolphins. Instead of using passive sound from other animals, these animals produce active sound into the environment and analyze the response to obtain localization or navigation information. Similarly, blind people can use their ears, experience and sound characteristics of an environment to locate themselves. An environment's sound field is perceived by animals or human beings through the phase differences and the magnitude ratio, called the interaural time difference (ITD) and the interaural level difference (ILD), respectively, among sound-receiving sensors. Many authors have explored the idea of using the ITD of two microphones for sound localization, by generalized cross-correlation (GCC)-based methods. However, the ILD is seldom applied for localization since it is considered unreliable [17], and lacks an explicit and stable relationship to be formulated by a straightforward algorithm [18]. Indeed, both ITD and ILD represent meaningful physical quantities for a sound field perception. It is the variation in a complicated sound field that makes them hard to be used by simple algorithms. Nakadai *et al.* proposed auditory epipolar geometry [19, 20] to extract the directional information of sound sources by the integration of ITD and IID at the position of the ears of a humanoid robot. To overcome the inaccuracy when sounds come from the periphery, they further proposed a method to model the humanoid head by scattering theory [21]. Instead of estimating the ITD and the ILD, this study proposes a Gaussian mixture-sound field landmark model (GM-SFLM) to model the ITD (the phase difference) and the ILD (the magnitude ratio) distributions of a sound source in different locations. The experimental results demonstrate that the GM-SFLM could accurately locate the robot in a non-stationary noisy indoor environment, and overcome the shelter effect and the microphone mismatch problem.

The proposed GM-SFLM is composed of the phase difference Gaussian mixture model (GMM) and the magnitude ratio GMM. These two GMMs are combined using proper weights and a method for determining the weights based on the measurement of confidence level is proposed to improve the location detection correct rate. Other parameters such as the mean, variance and mixture weights within each GMM are derived from a set of location-dependent and content-independent sound field data by maximizing the log-likelihood of the *a posteriori* probability using the expectation-maximization (EM) algorithm [22], which can guarantee a monotonic increase in the model's log-likelihood value. With this *a priori* GM-SFLM, the proposed system is able to localize robots in complex environments. Moreover, the proposed localization system does not depend on the geometric relationship between source locations and microphones, and can handle both near-field and far-field problems. The experiment indicates that when the robot is under the shelter effect, this system still provides high detection accuracy. Since only two uncalibrated microphones are needed, a PC with a stereo recording sound card can be employed to detect the robot's location, which may reduce the cost and power consumption of the system.

This paper is organized as follows. The following section discusses the related works. Section 3 presents the system architecture and the localization procedures. Section 4 describes the proposed GM-SFLM-based method in detail. Section 5 shows and discusses the experimental results. Conclusions are drawn in the final section.

2. RELATED WORKS

Audible range sound devices are relatively inexpensive and common to many mobile robots. Generally, sound devices (such as speakers) are used to generate the sound for robots to communicate with people, present robot emotions or alert users. If sound is allowed to be used for localization applications, then the hardware cost and implementation complexity is minimal. Furthermore, the audible range sound is omnidirectional, slowly fading and capable of transmitting a long distance in a complex enclosure. The concept of employing a microphone array to localize sound has been developed for over 30 years. The major procedures can be separated into three categories: steered-beamformer-based methods [23, 24], GCC methods [25, 26] and methods based on eigenstructure analysis [27]. Among these categories, GCC-based methods are the most appropriate for realizing with only two microphones. These methods estimate the time delay between microphone pairs and apply the sound propagation relation to obtain the source direction. The performance of conventional GCC-based methods is sensitive to the reverberation and noise. Brandstein *et al.* proposed Tukey's biweight to the weighting function to overcome the reflection effect [28].

In the robotic field, the methods mentioned above have also been broadly adopted. These methods can help the robot to locate the sound source of interest [29–33] or to localize the robot itself. Wang *et al.* proposed an acoustic robot navigation system [34] in which 24 microphones were separated into two linear arrays and mounted onto two orthogonal walls. The sound source's location (which is also the robot's location) is determined by maximizing the likelihood function constructed by the propagation relation and the GCC-based delay estimation. Owing to the array geometry, this system can estimate the robot's location in a two-dimensional environment. Valin *et al.* proposed a multiple microphone-based method with probabilistic post-processoring [35] to improve the robustness when some of the microphones are unable to receive the sound properly.

When only two microphones are used, the methods mentioned above estimate only the direction of the sound source, not the location. It means that the methods cannot distinguish between different sound sources that are aligned relative to the array. Furthermore, barriers may exist between the microphones and sound source (the so-called the shelter effect) in real applications. Under these circumstances, these methods estimate only the directions of reflection or diffraction and cannot determine the real source direction. In practice, microphone mismatch is also an important issue [36, 37], since the methods above assume that microphones are mutually matched. Pre-matched microphones are relatively expensive and the microphone calibration procedure is not always reliable because the characteristics of microphone change with sound direction and are hard to measure precisely.

3. SYSTEM IMPLEMENTATION

The experimental platform used in this work is a dog-like pet robot, which includes a quadruped robot (named '*e*Robot' hereafter) that can be transparently controlled through a wireless network. Figure 1 illustrates the localization scenario, where a robot localization agent is mounted in an arbitrarily indoor position. The *e*Robot can move and bark for model training and location detection.

Figure 2 depicts the *e*Robot, which has 16 d.o.f. in motion and an embedded Ethernet [38, 39] for distributed and parallel access of the actuators and sensors. A tiny network bridge integrates both wired and wireless networks (IEEE802.3 and IEEE802.11b) (see Fig. 3). Through this bridge, the actuators and sensors can be controlled and accessed transparently from any network-connected computer (e.g., the robot localization agent in Fig. 1). For detailed construction of the *e*Robot, refer

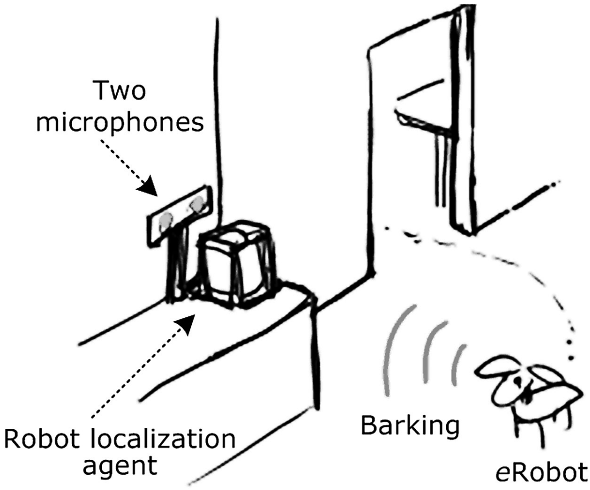


Figure 1. The overall system architecture.

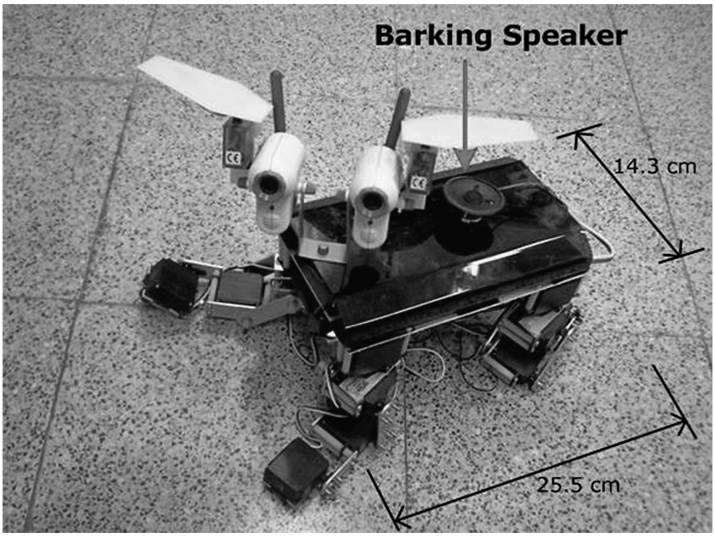


Figure 2. Photograph of *e*Robot (named *O-Di* robot).

to Refs [40, 47]. *e*Robot motion planning is performed using the two-wheel model [41] which includes commands such as forward, backward, turn around, etc.

The robot localization agent, which controls the *e*Robot to bark during localization, is realized on an x86-based PC. This agent contains two microphones and computes the GM-SFLM to locate the robot. Using the GM-SFLM, the robot localization agent can landmark, recollect and manage the *e*Robot's location by the barks from the *e*Robot. Figure 4 illustrates the robot localization methodology architecture, which can be separated into three stages.

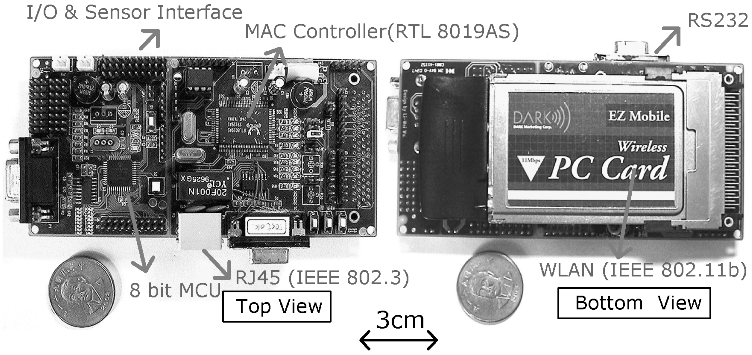


Figure 3. Photograph of tiny network bridge module.

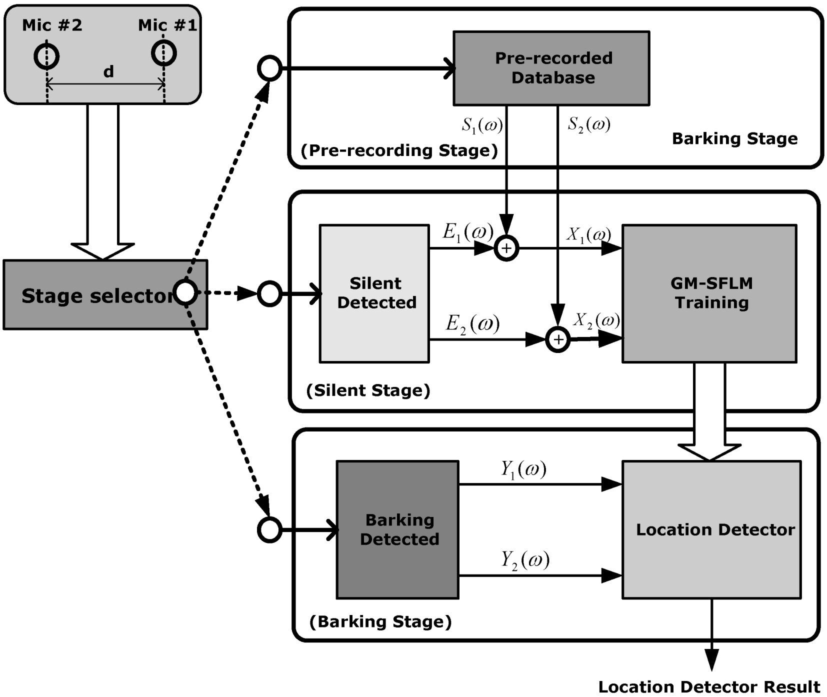


Figure 4. Robot localization methodology architecture.

3.1. First stage: pre-recording stage

In the first stage, called the pre-recording stage, the *e*Robot moves and barks in the locations of interest when the environment is quiet to obtain the pre-recorded database (denoted as $S_1(\omega)$ and $S_2(\omega)$ at each frequency ω from each microphone). The database is employed to acquire the sound field characteristic of each location.

3.2. Second stage: silent stage

In the second stage, called the silent stage (e.g., no barking), the environmental noise represented as $E_1(\omega)$ and $E_2(\omega)$ is recorded to collect the environmental noise characteristic. Assuming that noise is additive, the received signal can be expressed as a linear combination of barking signal and environmental noise. Under this assumption, the GM-SFLM is trained using signals $X_1(\omega) = S_1(\omega) + E_1(\omega)$ and $X_2(\omega) = S_2(\omega) + E_2(\omega)$. Details of the training procedures of the GM-SFLM are described in Section 4.

3.3. Third stage: barking stage

The third stage is the barking stage, in which the GM-SFLM is duplicated into the location detector to determine the robot's location. Since the testing sequence frame length is short in the barking stage, the noise characteristic is assumed to be the same as that in the silent stage. Hence, the GM-SFLM obtained in the silent stage can be adapted to the barking stage for location detection. Figure 5 shows the flowchart

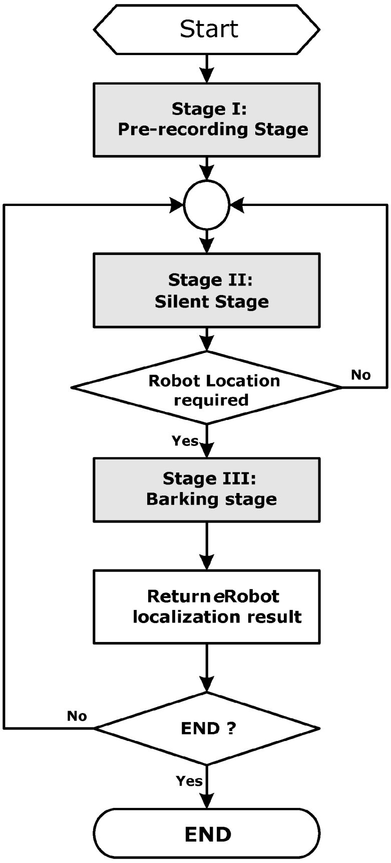


Figure 5. The flowchart of the robot localization system.

of the overall localization procedure. Additionally, a wireless Ethernet is adopted to accomplish the stage synchronization and communication between the robot and the robot localization agent.

4. ROBOT LOCALIZATION USING THE GM-SFLM

This section describes the proposed GM-SFLM, the procedures of obtaining the model parameters and the location detection algorithm.

4.1. GM-SFLM description

To establish a sound field landmark, the localization agent (Fig. 1) needs to construct models for the *e*Robot's barking sound at different locations. Generally, two microphones can provide the phase difference (i.e., time-delay) and the magnitude ratio information in the frequency domain expressed in:

$$M_x(\omega) = \frac{\|X_2(\omega)\|_2}{\|X_1(\omega)\|_2} \tag{1}$$

$$P_x(\omega) = \text{phase}(X_2(\omega)) - \text{phase}(X_1(\omega)), \tag{2}$$

where $\|\cdot\|_2$ denotes the two-norm operation. Theoretically, the phase and magnitude relate directly to the sound wave arrival direction and distance of the sound source. However, these simple relations only exist in free space or environments with simple geometry. In reality, complex boundary conditions and local sound scattering make these values impossible to use deterministically. Alternatively, in this work, the complexity enables the sound source to be located using the distributions of the statistical sound patterns at different source locations. GMMs [42] are introduced to model the distributions of the patterns. The sound pattern of a location contains the phase difference and the magnitude ratio information that are modeled using different GMMs, i.e., they are not joined together. Since the magnitude ratio distribution and the phase difference distribution depend only on the location of the sound source and are content-independent, they are easy to obtain whenever the *e*Robot is barking.

Denoting $P_x(\omega_b)$ and $M_x(\omega_b)$ as the phase difference and the magnitude ratio, respectively, at frequency $\omega_b, b = 1 - B$ and $\boldsymbol{P}_x = [P_x(\omega_1) \cdots P_x(\omega_B)]^{\mathrm{T}}$, $\boldsymbol{M}_x = [M_x(\omega_1) \cdots M_x(\omega_B)]^{\mathrm{T}}$ as the associated vectors. The GMMs are defined as the weighted sum of N_1 and N_2 mixtures of Gaussian component densities shown below:

$$G(\boldsymbol{P}_x|\boldsymbol{\lambda}_{\mathrm{P}}) = \sum_{i=1}^{N_1} \rho_{\mathrm{P},i}\, g_i(\boldsymbol{P}_x) \tag{3}$$

$$G(\boldsymbol{M}_x|\boldsymbol{\lambda}_{\mathrm{M}}) = \sum_{i=1}^{N_2} \rho_{\mathrm{M},i}\, g_i(\boldsymbol{M}_x), \tag{4}$$

where $\rho_{\mathrm{P},i}$ and $\rho_{\mathrm{M},i}$ are the i-th mixture weights, and $g_i(\boldsymbol{P}_x)$ and $g_i(\boldsymbol{M}_x)$ are the Gaussian density function defined later. The terms $\boldsymbol{\lambda}_\mathrm{P}$ and $\boldsymbol{\lambda}_\mathrm{M}$ in (3) and (4) represent the sets of mean vectors, covariance matrices and mixture weights from N_1 and N_2 component densities as:

$$\boldsymbol{\lambda}_\mathrm{P} = \{\boldsymbol{\rho}_\mathrm{P}, \boldsymbol{\mu}_\mathrm{P}, \boldsymbol{\Sigma}_\mathrm{P}\} \tag{5}$$

$$\boldsymbol{\lambda}_M = \{\boldsymbol{\rho}_\mathrm{M}, \boldsymbol{\mu}_\mathrm{M}, \boldsymbol{\Sigma}_\mathrm{M}\}, \tag{6}$$

where $\boldsymbol{\rho}_\mathrm{P} = [\boldsymbol{\rho}_{\mathrm{P},1} \cdots \boldsymbol{\rho}_{\mathrm{P},N_1}]$ denotes the phase difference mixture weight vector with dimensions $1 \times N_1$, $\boldsymbol{\rho}_\mathrm{M} = [\boldsymbol{\rho}_{\mathrm{M},1} \cdots \boldsymbol{\rho}_{\mathrm{M},N_2}]$ denotes the magnitude ratio mixture weight vector with dimensions $1 \times N_2$, $\boldsymbol{\mu}_\mathrm{P} = [\boldsymbol{\mu}_{\mathrm{P},1} \cdots \boldsymbol{\mu}_{\mathrm{P},N_1}]$ denotes the phase difference mean matrix with dimensions $B \times N_1$, $\boldsymbol{\mu}_\mathrm{M} = [\boldsymbol{\mu}_{\mathrm{M},1} \cdots \boldsymbol{\mu}_{\mathrm{M},N_2}]$ denotes the magnitude ratio mean matrix with dimensions $B \times N_2$, $\boldsymbol{\Sigma}_\mathrm{P} = [\boldsymbol{\Sigma}_{\mathrm{P},1} \cdots \boldsymbol{\Sigma}_{\mathrm{P},N_1}]$ denotes the phase difference covariance matrix with dimensions $B \times BN_1$ and $\boldsymbol{\Sigma}_\mathrm{M} = [\boldsymbol{\Sigma}_{M,1} \cdots \boldsymbol{\Sigma}_{\mathrm{M},N_2}]$ denotes the magnitude ratio covariance matrix with dimensions $B \times BN_2$.

The corresponding vectors and matrices defined above are:

$$\boldsymbol{\mu}_{\mathrm{P},i} = [\mu_{\mathrm{P},i}(\omega_1) \cdots \mu_{\mathrm{P},i}(\omega_B)]^\mathrm{T} \quad \text{and} \quad \boldsymbol{\mu}_{\mathrm{M},i} = [\mu_{\mathrm{M},i}(\omega_1) \cdots \mu_{\mathrm{M},i}(\omega_B)]^\mathrm{T}$$

$$\boldsymbol{\Sigma}_{\mathrm{P},i} = \begin{bmatrix} \sigma^2_{\mathrm{P},i}(\omega_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\mathrm{P},i}(\omega_B) \end{bmatrix} \quad \text{and}$$

$$\boldsymbol{\Sigma}_{\mathrm{M},i} = \left[\begin{pmatrix} \sigma^2_{\mathrm{M},i}(\omega_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\mathrm{M},i}(\omega_B) \end{pmatrix}\right].$$

Both $g_i(\boldsymbol{P}_x)$ and $g_i(\boldsymbol{M}_x)$ in (3) and (4) can now be given using these notations as:

$$g_i(\boldsymbol{P}_x) = \frac{1}{(2\pi)^{B/2}|\boldsymbol{\Sigma}_{\mathrm{P},i}|^2} \exp\left(-\frac{1}{2}[\boldsymbol{P}_x - \boldsymbol{\mu}_{\mathrm{P},i}]^\mathrm{T}\boldsymbol{\Sigma}^{-1}_{\mathrm{P},i}[\boldsymbol{P}_x - \boldsymbol{\mu}_{\mathrm{P},i}]\right) \tag{7}$$

$$g_i(\boldsymbol{M}_x) = \frac{1}{(2\pi)^{B/2}|\boldsymbol{\Sigma}_{\mathrm{M},i}|^2} \exp\left(-\frac{1}{2}[\boldsymbol{M}_x - \boldsymbol{\mu}_{\mathrm{M},i}]^\mathrm{T}\boldsymbol{\Sigma}^{-1}_{\mathrm{M},i}[\boldsymbol{M}_x - \boldsymbol{\mu}_{\mathrm{M},i}]\right). \tag{8}$$

Notably, the mixture weights in (3) and (4) must satisfy the constraints:

$$\sum_{i=1}^{N_1} \rho_{\mathrm{P},i} = 1 \quad \text{and} \quad \sum_{i=1}^{N_2} \rho_{\mathrm{M},i} = 1. \tag{9}$$

The proposed GM-SFLM at each location is defined as the linear combination of the phase difference GMM and the magnitude ratio GMM as:

$$F_{\mathrm{GM-SFLM}} = \alpha_\mathrm{P} G(\boldsymbol{P}_x|\boldsymbol{\lambda}_\mathrm{P}) + \alpha_\mathrm{M} G(\boldsymbol{M}_x|\boldsymbol{\lambda}_\mathrm{M}), \tag{10}$$

where α_p and α_M represent the weighting factors. The covariance matrices, $\boldsymbol{\Sigma}_\mathrm{P}$ and $\boldsymbol{\Sigma}_\mathrm{M}$, are selected as diagonal matrices. Although both the phase difference and the magnitude ratio between microphone pairs may not be statistically independent, the GMM with diagonal covariance matrices can model the data correlations by increasing the mixture number [43].

4.2. GM-SFLM parameters estimation

The parameters $\boldsymbol{\lambda}_{\mathrm{P}}$ and $\boldsymbol{\lambda}_{\mathrm{M}}$ in (5) and (6) can be estimated by the maximum likelihood (ML) approach, which estimates the model parameters by maximizing the log-likelihood of the GMMs as:

$$\log G(\boldsymbol{P}_x|\boldsymbol{\lambda}_{\mathrm{P}}) = \sum_{t=1}^{T} \log G(\boldsymbol{P}_x^{(t)}|\boldsymbol{\lambda}_{\mathrm{P}}) \tag{11}$$

$$\log G(\boldsymbol{M}_x|\boldsymbol{\lambda}_{M}) = \sum_{t=1}^{T} \log G(\boldsymbol{M}_x^{(t)}|\boldsymbol{\lambda}_{\mathrm{M}}), \tag{12}$$

where the superscript (t) denotes the t-th frame, and $\mathbf{P}_x = \{\boldsymbol{P}_x^{(1)}, \ldots, P_x^{(T)}\}$ and $\mathbf{M}_x = \{\boldsymbol{M}_x^{(1)}, \ldots, \boldsymbol{M}_x^{(T)}\}$ are the sequences of the T-th input feature vectors. However, direct maximization of the log-likelihood function for the mixture models is numerically difficult due to their nonlinearity and strong coupling. This work applies the iterative EM algorithm [22] which can guarantee a monotonic increase in the model's log-likelihood value. The iterative procedure can be separated into the following two steps:
Expectation step:

$$G(i|\boldsymbol{P}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{P}}) = \frac{\rho_{\mathrm{P},i}\, g_i(\boldsymbol{P}_x^{(t)})}{\sum_{i=1}^{N_1} \rho_{\mathrm{P},i}\, g_i(\boldsymbol{P}_x^{(t)})} \tag{13}$$

$$G(i|\boldsymbol{M}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{M}}) = \frac{\rho_{M,i}\, g_i(M_x^{(t)})}{\sum_{i=1}^{N_2} \rho_{\mathrm{M},i}\, g_i(\boldsymbol{M}_x^{(t)})}, \tag{14}$$

where $G(i|\boldsymbol{P}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{P}})$ and $G(i|\boldsymbol{M}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{M}})$ are *a posteriori* probabilities.

(i) Estimate the mixture weights:

$$\rho_{\mathrm{P},i} = \frac{1}{T} \sum_{t=1}^{T} G(i|\boldsymbol{P}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{P}}) \tag{15}$$

$$\rho_{\mathrm{M},i} = \frac{1}{T} \sum_{t=1}^{T} G(i|\boldsymbol{M}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{M}}). \tag{16}$$

(ii) Estimate the mean vector:

$$\boldsymbol{\mu}_{\mathrm{P},i} = \frac{\sum_{t=1}^{T} G(i|\boldsymbol{P}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{P}})\boldsymbol{P}_x^{(t)}}{\sum_{t=1}^{T} G(i|\boldsymbol{P}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{P}})} \tag{17}$$

$$\boldsymbol{\mu}_{\mathrm{M},i} = \frac{\sum_{t=1}^{T} G(i|\boldsymbol{M}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{M}})\boldsymbol{M}_x^{(t)}}{\sum_{t=1}^{T} G(i|\boldsymbol{M}_x^{(t)}, \boldsymbol{\lambda}_{\mathrm{M}})}. \tag{18}$$

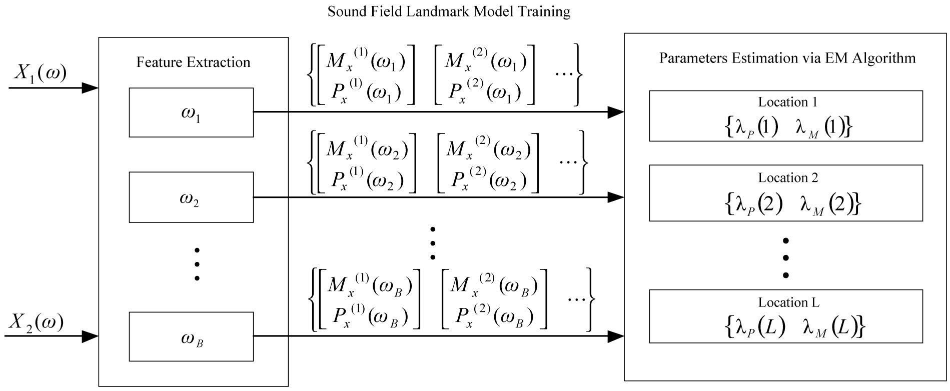

Sound Field Landmark Model Training
$X_1(\omega)$
$X_2(\omega)$
Feature Extraction
ω_1
ω_2
ω_B
$\left\{\begin{bmatrix} M_x^{(1)}(\omega_1) \\ P_x^{(1)}(\omega_1) \end{bmatrix} \begin{bmatrix} M_x^{(2)}(\omega_1) \\ P_x^{(2)}(\omega_1) \end{bmatrix} \cdots\right\}$
$\left\{\begin{bmatrix} M_x^{(1)}(\omega_2) \\ P_x^{(1)}(\omega_2) \end{bmatrix} \begin{bmatrix} M_x^{(2)}(\omega_2) \\ P_x^{(2)}(\omega_2) \end{bmatrix} \cdots\right\}$
$\left\{\begin{bmatrix} M_x^{(1)}(\omega_B) \\ P_x^{(1)}(\omega_B) \end{bmatrix} \begin{bmatrix} M_x^{(2)}(\omega_B) \\ P_x^{(2)}(\omega_B) \end{bmatrix} \cdots\right\}$
Parameters Estimation via EM Algorithm
Location 1
$\{\lambda_P(1)\ \lambda_M(1)\}$
Location 2
$\{\lambda_P(2)\ \lambda_M(2)\}$
Location L
$\{\lambda_P(L)\ \lambda_M(L)\}$

Figure 6. SFLM training procedure.

(iii) Estimate the variances:

$$\sigma^2_{\mathrm{P},i}(\omega_b) = \frac{\sum_{t=1}^{T} G(i|\boldsymbol{P}_x^{(t)}, \boldsymbol{\lambda}_\mathrm{P}) P_x^{(t)2}(\omega_b)}{\sum_{t=1}^{T} G(i|\boldsymbol{P}_x^{(t)}, \boldsymbol{\lambda}_\mathrm{P})} - \mu^2_{\mathrm{P},i}(\omega_b) \tag{19}$$

$$\sigma^2_{\mathrm{M},i}(\omega_b) = \frac{\sum_{t=1}^{T} G(i|\boldsymbol{M}_x^{(t)}, \boldsymbol{\lambda}_\mathrm{M}) M_x^{(t)2}(\omega_b)}{\sum_{t=1}^{T} G(i|\boldsymbol{M}_x^{(t)}, \boldsymbol{\lambda}_\mathrm{M})} - \mu^2_{\mathrm{M},i}(\omega_b), \tag{20}$$

where $b = \{1, \ldots, B\}$.

However, the EM algorithm only guarantees to find a local maximum log-likelihood model which is sensitive to the choice of initial model. K-means [43] is by far the most widely used method to obtain the initial model. Charles [44] proposed an accelerated K-means algorithm, which utilizes the triangle inequality to significantly reduce the computational power requirement. Charles' method is also suitable for discovering an appropriate initial model to lower the iteration number of the EM algorithm. Figure 6 depicts the location model training procedure with the total location number L.

The values of α_P and α_M can be chosen arbitrarily. However, poor choices of these parameters would lead to a poor localization result. This work provides a method to determine these parameters based on the sum of the correlation values among locations of the phase difference GMM and magnitude ratio GMM. The GMM with the higher correlation value sum would be assigned a low weight, since the ability to discriminate is considered lower under this circumstance, and *vice versa*. Under this principle, α_P and α_M are determined by the following formula:

$$\begin{aligned} &\min\left\{\sum_{q_p} \alpha_\mathrm{p}\{\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})\mathbf{U}\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})^\mathrm{T}\} + \sum_{\boldsymbol{q}_\mathrm{M}} \alpha_\mathrm{M}\{\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})\mathbf{U}\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})^\mathrm{T}\}\right\} \\ &\text{s.t.}\quad \alpha_\mathrm{P}\alpha_\mathrm{M} = 1, \alpha_\mathrm{P} > 0, \alpha_\mathrm{M} > 0, \end{aligned} \tag{21}$$

where $\boldsymbol{q}_{\mathrm{P}} \in Q_{\mathrm{p}}$ and $\boldsymbol{q}_{\mathrm{M}} \in Q_{\mathrm{M}}$ are the B-dimensional random vectors in the operation ranges, Q_{p} and Q_{M}:

$$\begin{aligned}\mathbf{C}_{\mathrm{P}}(\boldsymbol{q}_{\mathrm{P}}) &= \begin{bmatrix} C(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(1)) & C(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(2)) & \cdots & C(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(L)) \end{bmatrix},\\ \mathbf{C}_{\mathrm{M}}(\boldsymbol{q}_{\mathrm{M}}) &= \begin{bmatrix} C(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(1)) & C(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(2)) & \cdots & C(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(L)) \end{bmatrix},\end{aligned}$$

and:

$$\mathbf{U} = \begin{bmatrix} 0 & 1 & 1 & \cdots & \cdots & 1 \\ 0 & 0 & 1 & 1 & \cdots & 1 \\ \vdots & 0 & 0 & \ddots & \cdots & 1 \\ \vdots & \vdots & 0 & \ddots & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{with dimension } L \times L.$$

In addition:

$$\begin{aligned} C(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(l)) &= \frac{H(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(l))}{\sqrt{\sum_{\boldsymbol{q}_{\mathrm{p}}} H^2(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(l))}} \\ C(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(l)) &= \frac{H(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(l))}{\sqrt{\sum_{\boldsymbol{q}_{\mathrm{M}}} H^2(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(l))}}, \\ H(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(l)) &= G(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(l)) - \frac{\sum_{\boldsymbol{q}_{\mathrm{p}}} G(\boldsymbol{q}_{\mathrm{P}}|\boldsymbol{\lambda}_{\mathrm{P}}(l))}{N(\boldsymbol{q}_{\mathrm{P}})}, \end{aligned}$$

and:

$$H(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(l)) = G(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(l)) - \frac{\sum_{\boldsymbol{q}_{\mathrm{M}}} G(\boldsymbol{q}_{\mathrm{M}}|\boldsymbol{\lambda}_{\mathrm{M}}(l))}{N(\boldsymbol{q}_{\mathrm{M}})},$$

where $N(\boldsymbol{q}_{\mathrm{P}})$ and $N(\boldsymbol{q}_{\mathrm{M}})$ denote the total selected numbers of $\boldsymbol{q}_{\mathrm{P}}$ and $\boldsymbol{q}_{\mathrm{M}}$.

The values of α_{P} and α_{M} can be obtained by solving (21). The proof is given in the Appendix:

$$\alpha_{\mathrm{P}} = \sqrt{\frac{\sum_{\boldsymbol{q}_{\mathrm{M}}} \mathbf{C}_{\mathrm{M}}(\boldsymbol{q}_{\mathrm{M}})\mathbf{U}\mathbf{C}_{\mathrm{M}}(\boldsymbol{q}_{\mathrm{M}})^{\mathrm{T}}}{\sum_{\boldsymbol{q}_{\mathrm{p}}} \mathbf{C}_{\mathrm{P}}(\boldsymbol{q}_{\mathrm{P}})\mathbf{U}\mathbf{C}_{\mathrm{P}}(\boldsymbol{q}_{\mathrm{P}})^{\mathrm{T}}}} \tag{22}$$

$$\alpha_{\mathrm{M}} = \sqrt{\frac{\sum_{\boldsymbol{q}_{\mathrm{p}}} \mathbf{C}_{\mathrm{P}}(\boldsymbol{q}_{\mathrm{P}})\mathbf{U}\mathbf{C}_{\mathrm{P}}(\boldsymbol{q}_{\mathrm{P}})^{\mathrm{T}}}{\sum_{\boldsymbol{q}_{\mathrm{M}}} \mathbf{C}_{\mathrm{M}}(\boldsymbol{q}_{\mathrm{M}})\mathbf{U}\mathbf{C}_{\mathrm{M}}(\boldsymbol{q}_{\mathrm{M}})^{\mathrm{T}}}}. \tag{23}$$

4.3. Location detection

The parameters $\boldsymbol{\lambda}_{\mathrm{P}}(1), \ldots, \boldsymbol{\lambda}_{\mathrm{P}}(L), \boldsymbol{\lambda}_{\mathrm{M}}(1), \ldots, \boldsymbol{\lambda}_{\mathrm{M}}(L)$, α_{P} and α_{M} represent the GM-SFLM of L locations, and the location is determined by finding the maximum *a posteriori* probability for a given observation sequence:

$$\begin{aligned}\hat{l} &= \arg\max_{1\leqslant l\leqslant L} F_{\mathrm{GM-SFLM}}(l) = \arg\max_{1\leqslant l\leqslant L} \alpha_{\mathrm{P}} G(\boldsymbol{\lambda}_{\mathrm{P}}(l)|\mathbf{P}_{\mathrm{Y}}) + \alpha_{\mathrm{M}} G(\boldsymbol{\lambda}_{\mathrm{M}}(l)|\mathbf{M}_{\mathrm{Y}}) \\ &= \arg\max_{1\leqslant l\leqslant L} \alpha_{\mathrm{P}} \frac{G(\mathbf{P}_{\mathrm{Y}}|\boldsymbol{\lambda}_{\mathrm{P}}(l))p(\boldsymbol{\lambda}_{\mathrm{P}}(l))}{p(\mathbf{P}_{\mathrm{Y}})} + \alpha_{M} \frac{G(\mathbf{M}_{\mathrm{Y}}|\boldsymbol{\lambda}_{\mathrm{M}}(l))p(\boldsymbol{\lambda}_{\mathrm{M}}(l))}{p(\mathbf{M}_{\mathrm{Y}})},\end{aligned} \tag{24}$$

where $\mathbf{P}_Y = \{\boldsymbol{P}_Y^{(1)}, \ldots, \boldsymbol{P}_Y^{(V)}\}$ and $\mathbf{M}_Y = \{\boldsymbol{M}_Y^{(1)}, \ldots, \boldsymbol{M}_Y^{(V)}\}$ are the phase difference and the magnitude ratio computed from the testing sequences denoted as $Y_1(\omega)$ and $Y_2(\omega)$ and illustrated in Fig. 4, and V denotes the testing sequence length. The probabilities $p(\boldsymbol{\lambda}_{\mathrm{P}}(l))$ and $p(\boldsymbol{\lambda}_{\mathrm{M}}(l))$ could be selected as $1/L$ since the probability in each position is equally likely for a blind search. Since the probability densities $p(\mathbf{P}_Y)$ and $p(\mathbf{M}_Y)$ are the same for all position models, the detection rule can be recast as:

$$\hat{l} = \arg\max_{1\leqslant l\leqslant L} \alpha_{\mathrm{P}} \prod_{v=1}^{V} G\big(\boldsymbol{P}_Y^{(v)}|\boldsymbol{\lambda}_{\mathrm{P}}(l)\big) + \alpha_{\mathrm{M}} \prod_{v=1}^{V} G\big(\boldsymbol{M}_Y^{(v)}|\boldsymbol{\lambda}_{\mathrm{M}}(l)\big). \tag{25}$$

5. EXPERIMENTAL RESULTS

The first experiment was conducted using two microphones in a small room as shown in Fig. 7. A total of 12 locations were defined for the test. The barking signal of *e*Robot is illustrated in Fig. 8. Since the major frequencies of the barking signal were limited to 1.7 kHz, the two microphones were spaced 0.1 m apart to avoid the spatial aliasing effect [45]. In this experiment, the same barking sounds are used in the first and the third stage. However, it is not necessary to restrict the sounds produced by the robot as long as they have similar major frequencies because the proposed features of phase difference and magnitude ratio distributions are content independent. Considering the size of the *e*Robot, the location blocks were assigned a radius of 0.4 m. The shelter effect occurred in this experimental environment when the robot barked in the partitioned room, e.g., locations 1, 2 and 3 in Fig. 7. Additionally, locations 4 and 5 do not have direct sound paths to the two microphones because of the table barrier. Figure 9 depicts the relative physical configuration of the experimental environment and the robot. The experiment was performed under four different signal-to-noise (SNR) conditions—the first in a quiet environment and the others with background speech. Table 1 lists the SNR ranges of the four cases. The received signals were sampled at 8 kHz and the window for the short time Fourier transform (STFT) contained 256 zero padding samples and 32-ms speech signals, totaling 512 samples. Figure 10 shows the processed frame and the overlapping condition.

The covariance values update in (19) and (20) may lead to numerical difficulties, as the covariance matrices become nearly singular. In the experiment, the lower bounds of the variances $\sigma_{\mathrm{P},i}^2$ and $\sigma_{\mathrm{M},i}^2$ were set to 0.02 and 0.01, respectively. The selected frequency number B was set as 6, the training frame number was 500,

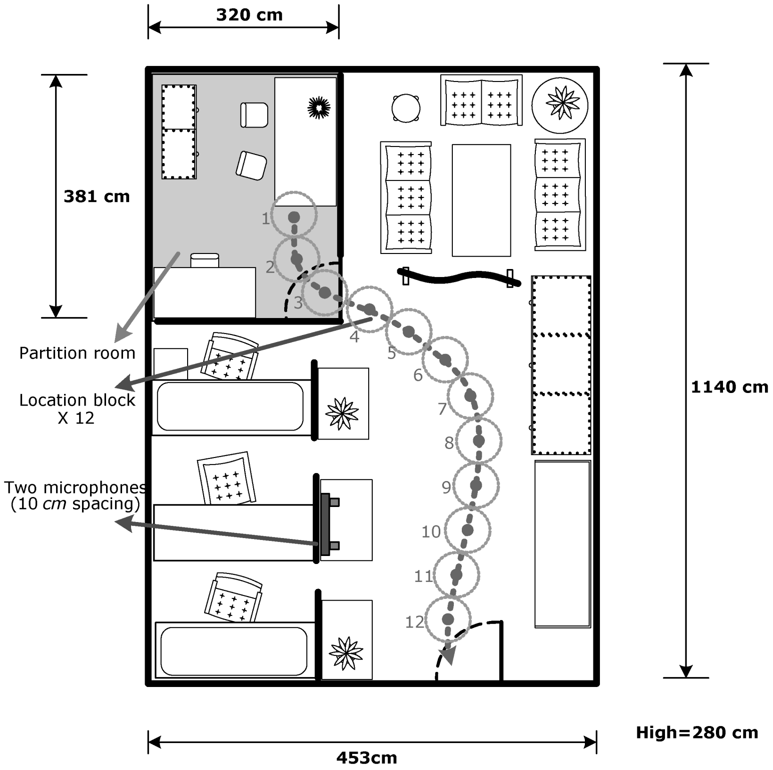


Figure 7. Room map of the robot location experiment (the environment is complex and includes a partition room).

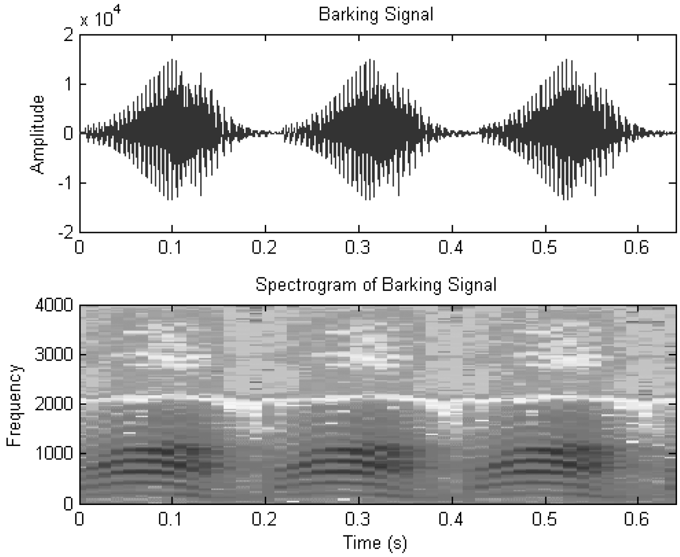


Figure 8. The waveform and spectrogram of the barking signal.

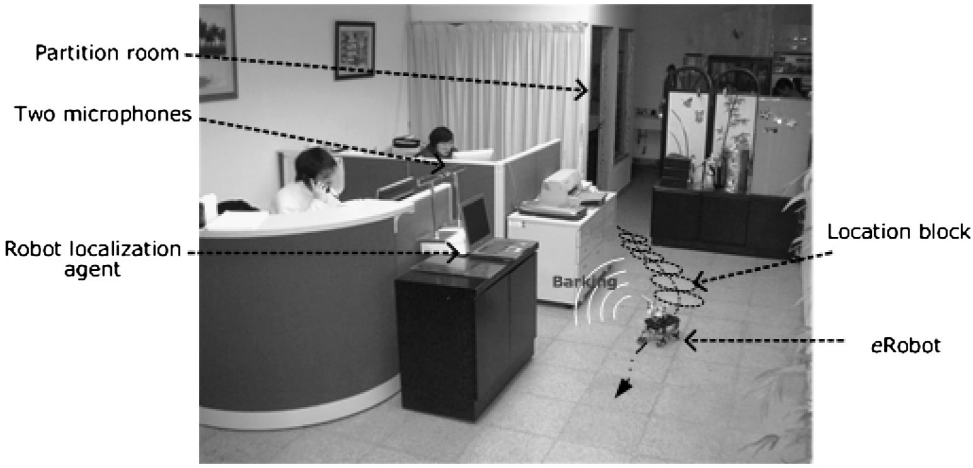


Figure 9. The relative physical configuration of the experimental environment and the *e*Robot.

Table 1.
The SNR ranges of the four different conditions

Conditions	Measured SNR ranges (dB)
1	– (quiet)
2	12.30–16.79
3	5.41–10.22
4	−12.89 to −8.71

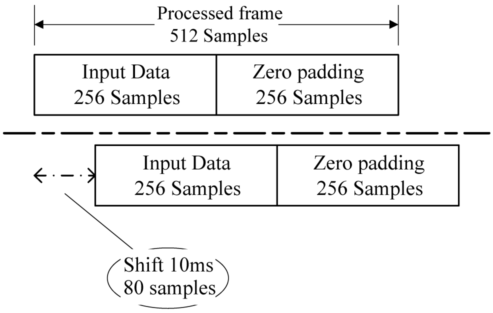
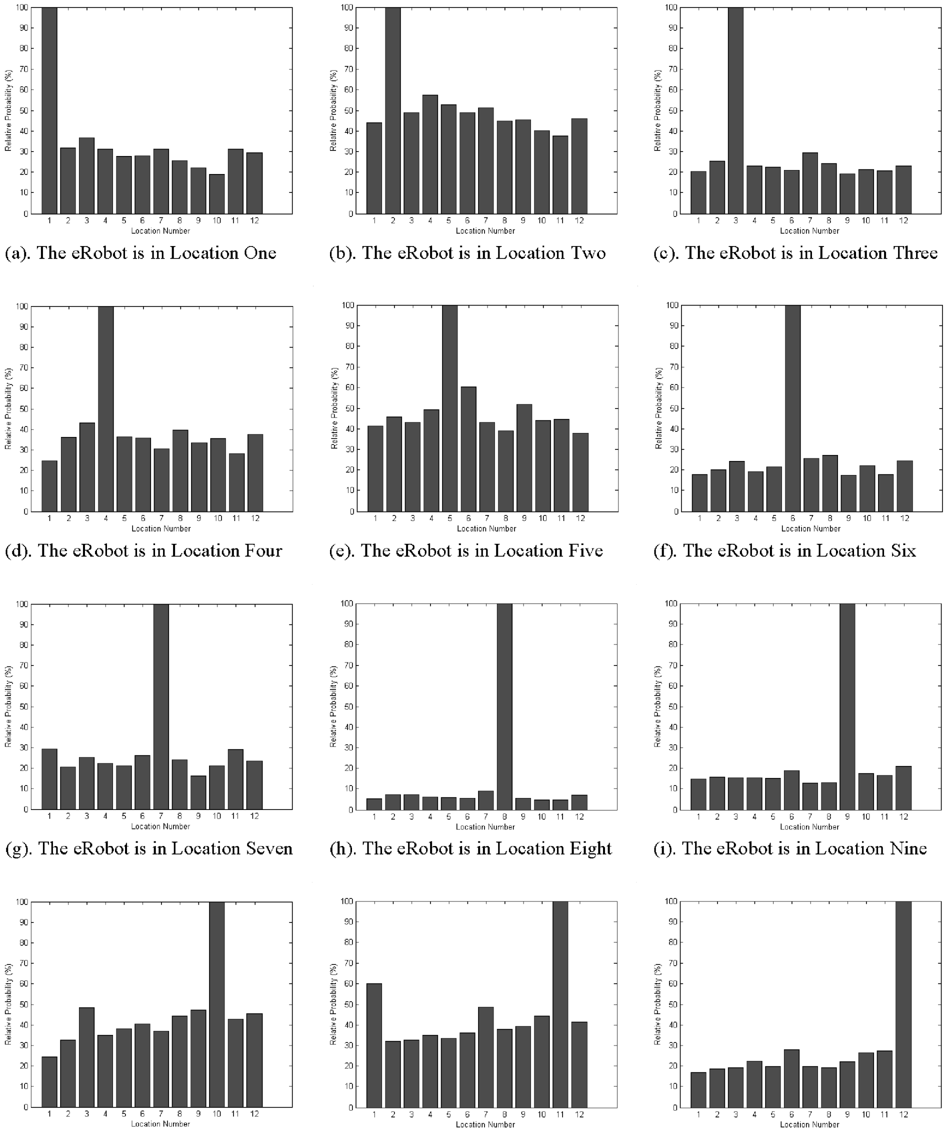


Figure 10. A processed frame and overlapping condition.

and the frame length of the training sequence T and the testing sequence Vwere set to 200 and 20, respectively. In other words, a 5-s barking sound recorded in the pre-recording stage was used in the silent stage for training and a 200-ms sound was used in the barking stage for testing (see Fig. 4). Additionally, five different mixture numbers of GM-SFLM, 1, 2, 4, 6 and 8, were used to evaluate the performance. Nine values separated uniformly in the operation ranges were specified for each element of $\boldsymbol{q}_{\mathrm{P}}$ and $\boldsymbol{q}_{\mathrm{M}}$ to set the values of α_{P} and α_{M} using (22) and (23). Figure 11



(a). The eRobot is in Location One

(b). The eRobot is in Location Two

(c). The eRobot is in Location Three

(d). The eRobot is in Location Four

(e). The eRobot is in Location Five

(f). The eRobot is in Location Six

(g). The eRobot is in Location Seven

(h). The eRobot is in Location Eight

(i). The eRobot is in Location Nine

(j). The eRobot is in Location Ten

(k). The eRobot is in Location Eleven

(l). The eRobot is in Location Twelve

Figure 11. The relative probability of 12 locations.

shows the relative probability distributions (RPD) of one trial using the mixture number of 8. The RPD is defined as

$$\mathrm{RPD}(l) = \frac{\alpha_{\mathrm{P}} G(\boldsymbol{\lambda}_{\mathrm{P}}(l)|\mathbf{P}_Y) + \alpha_{\mathrm{M}} G(\boldsymbol{\lambda}_{\mathrm{M}}(l)|\mathbf{M}_Y)}{\max_{1\leqslant l\leqslant L}[\alpha_{\mathrm{P}} G(\boldsymbol{\lambda}_{\mathrm{P}}(l)|\mathbf{P}_Y) + \alpha_{\mathrm{M}} G(\boldsymbol{\lambda}_{\mathrm{M}}(l)|\mathbf{M}_Y)]} \times 100\%. \qquad (26)$$

The experiment is performed when the robot stops and the experimental results indicate that the RPD of the correct *e*Robot location is distinctly separated from the other locations and shows that the proposed method could easily detect the location.

Tables 2–5 show the correct rates for each experimental condition (Table 1) with different mixture numbers in the GM-SFLM. The trial number of each location is 100.

Notably, the experiment was performed blindly, without prior knowledge of the robot's position and heading direction. The analytical results demonstrate that the correct rate is high especially when the mixture number is properly selected and indicate that the sound field characteristic can be captured by modeling its

Table 2.
Experimental result (%) in condition 1 using the GM-SFLM

Mixture number	Location number											
	1	2	3	4	5	6	7	8	9	10	11	12
1	99	69	31	99	25	100	100	100	83	100	100	100
2	100	100	96	100	63	100	100	100	100	100	100	100
4	100	100	100	100	97	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100	100	100	100	100	100

Table 3.
Experimental result (%) in condition 2 using the GM-SFLM

Mixture number	Location number											
	1	2	3	4	5	6	7	8	9	10	11	12
1	100	72	41	95	19	100	96	77	95	72	75	82
2	100	99	89	100	90	100	100	100	99	99	98	100
4	100	99	100	100	48	100	100	100	100	100	94	100
6	100	100	100	100	100	100	100	100	100	100	97	100
8	100	100	100	100	100	100	100	100	100	100	100	100

Table 4.
Experimental result (%) in condition 3 using the GM-SFLM

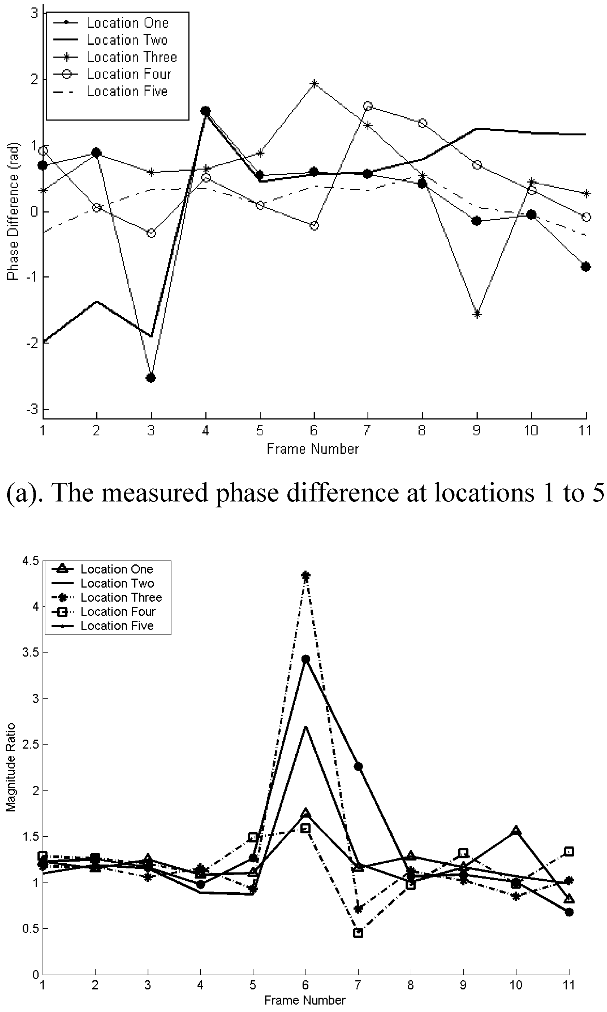
Mixture number	Location number											
	1	2	3	4	5	6	7	8	9	10	11	12
1	99	37	69	86	8	100	95	100	98	93	84	81
2	100	98	96	94	83	100	97	100	99	95	92	94
4	100	99	100	100	92	100	100	100	100	99	91	100
6	100	99	100	99	95	100	100	100	100	100	94	100
8	100	100	100	100	96	100	100	100	100	100	98	100

Table 5.
Experimental result (%) in condition 4 using the GM-SFLM

Mixture number	Location number											
	1	2	3	4	5	6	7	8	9	10	11	12
1	76	10	21	31	5	32	74	67	72	75	81	42
2	97	62	22	78	74	75	78	79	95	79	82	74
4	100	99	93	94	86	89	98	100	96	94	90	88
6	100	100	94	94	96	97	100	100	100	97	97	94
8	100	100	100	100	98	100	100	100	100	98	98	99

complexity. The correct rate between Locations 1 and 2 is also interesting. When the mixture number is 1 (i.e., using a simple Gaussian distribution), the correct rate at Location 1 is much higher than it is at Location 2. However, these two locations are close to each other. Significantly, both the phase difference and the magnitude ratio modeled by the GM-SFLM are physical quantities from wave propagation, rather than artificially formed variables. This observation demonstrates that under the shelter effect, the sound field characteristic changes drastically even with a small movement in space of the source. Location 5 exhibits similar results when compared with Locations 4 and 6.

To demonstrate that the locations cannot be found using the sound arrival delay between the microphones, Fig. 12 plots the phase differences and magnitude ratios at 0.65625 kHz at Locations 1 to 5. The data in Fig. 12 were obtained when the robot was motionless and the environment was quiet. Note that all the selected frames contained the barking sound and a local spectral peak of 0.65625 kHz. The plot shows that the values do not correspond accurately to the sound arrival direction. In brief, the phase differences of Locations 1–5 for the non-line-of-sight cases are not constant to represent the accurate arrival direction, so does the magnitude ratio. Furthermore, the overlapped variation of the values makes it hard to determine the range for location detection. Nevertheless, the proposed approach provides an accurate and robust location detection result in a complex environment. Another interesting phenomenon is the values of α_P and α_M, which are determined by (22) and (23). As the values of α_P and α_M vary with mixture numbers, GMMs of the modeled locations, and SNR conditions, it is hard to show all the values of α_P and α_M under all conditions. Generally, it is believed that the ITD dominates at lower frequency and the ILD dominates at higher frequency. However, the values of α_P are not always larger or smaller than the values of α_M. This is because the values of α_P and α_M depend on the correlation values among locations instead of the traditional physical concept of IID and IPD, and the phase difference distribution at lower frequency may have a higher correlation than the magnitude ratio distribution, which means that α_P is smaller than α_M. Consequently, we are unable to determine which distribution would dominate. This work demonstrates that the sound field can be applied for localization when the complexity is accurately modeled.



(a). The measured phase difference at locations 1 to 5

(b). The measured magnitude ratio at locations 1 to 5

Figure 12. The measured phase difference and magnitude ratio at Locations 1–5.

The second experiment was performed to demonstrate the location detection result when the robot was moving on a line from the robot localization agent. The configuration of the experiment is illustrated in Fig. 13. Table 6 shows the correct rates of each location when the mixture number is 8 and the trial number of each location is 60. As the locations are aligned to the microphones, it is hard to discriminate them by using traditional methods. However, as shown in Table 6, the correct rates of the proposed method still remain more than 80% even under low SNR conditions.

A well-known eigenstructure-based DOA estimator named MUSIC [27] with six microphones was used to compare with the proposed method. The distance between

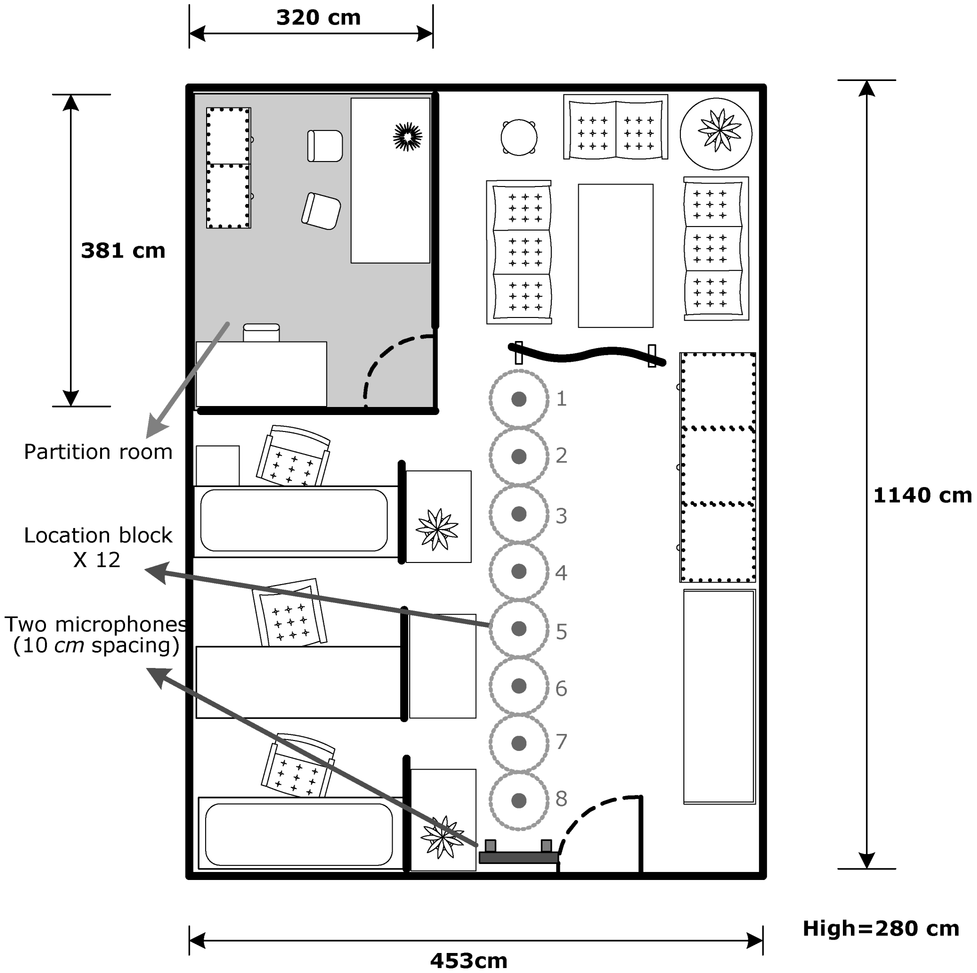


Figure 13. Room map of the robot location experiment (the environment is complex and includes a partition room).

Table 6.
Correct rate (%) using the GM-SFLM when the robot is on a line from the robot localization agent

Condition	Location number							
	1	2	3	4	5	6	7	8
1	98.33	100	97	98	100	100	98	100
2	97	100	99	100	100	99	100	100
3	97	100	97	98	100	100	97	98
4	83	100	95	98	98	100	92	88

Table 7.
Experimental results (%) of the MUSIC algorithm

Condition	Location number						
	6	7	8	9	10	11	12
1	100	55	84	83	87	68	59
2	100	56	84	83	86	69	58
3	100	56	85	83	86	69	57
4	99	48	87	76	78	62	41

microphones was 0.1 m. The experimental conditions were the same as the proposed method and the arithmetic mean-based frequency band combination was adopted. The frequency bands selected were those of the proposed method. Locations 1–5 were not taken into consideration in the experiment because they were under the non-line-of-sight condition. Since the proposed method utilizes *a priori* information of sound field characteristics, a non-blind the MUSIC approach [46] that measures the transfer functions from the possible locations to the microphones is adopted. Table 7 shows the correct rates of the localization results based on the MUSIC method. As shown in Table 7, the MUSIC-based method could not provide satisfactory correct rates, especially under low SNR cases.

6. CONCLUSIONS

This work proposes a robust robot localization system based on the sound field using only two microphones in an indoor environment. The proposed method can overcome practical issues such as the microphone's mismatch, near-field effect, shelter effect, and problems common in complex environments such as scattering and coherent reflection. The key point is that these issues make the sound field sophisticated enough to be discriminated with the proposed method if properly modeled. This work proves that the proposed method can capture the sound field characteristic with a very high localization correct rate. The accurate and robust experimental results indicate a promising direction of using sound as a means of localization where the devices are relatively inexpensive. However, several issues can be explored further, such as the relation of the proposed model to acoustic scattering theory and three-dimensional landmarks. These areas will be the work of continuing research by the authors.

Acknowledgments

The authors would like to thank the National Science Council of the Republic of China, Taiwan (contract no. NSC 92-2213-E-009-004) and the Ministry of Education of the Republic of China, Taiwan (contract no. 91-E-FA06-4-4) for financially supporting this research.

REFERENCES

1. H. R. Everrett *Sensors for Mobile Robots: Theory and Application*. Peters, Wellesley, MA (1995).
2. J. Borenstein, H. R. Everett and L. Feng, *Navigating Mobile Robots: Sensors and Techniques*. Peters, Wellesley, MA (1996).
3. T. Makimoto and Y. Sakai, Evolution of low power electronics and its future applications, in: *Proc. Int. Symp. on Low Power Electronics and Design*, New York, pp. 2–5 (2003).
4. A. Georgiev and P. K. Allen, Localization methods for a mobile robot in urban environments, *IEEE Trans. Robotics* **20**, 851–864 (2004).
5. S. Ghidary, T. Tani, T. Takamori and M. Hattori, A new home robot positioning system (HRPS) using IR switched multi ultrasonic sensors, in: *Proc. IEEE Int. Conf. on SMC*, Tokyo, pp. 737–741 (1999).
6. C. D. McGillem and T. S. Rappaport, Infra-red location system for navigation of autonomous vehicles, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, Philadelphia, PA, pp. 1236–1238 (1988).
7. G. Borghi and V. Caglioti, Minimum uncertainty explorations in the self-localization of mobile robots, *IEEE Trans. Robotics and Automation* **14**, 902–911 (1998).
8. D. G. Morgenthaler, S. Hennessy, and D. DeMenthon, Range–video fusion and comparison of inverse perspective algorithms in static images, *IEEE Trans. Syst., Man Cybernet.* **20**, 1301–1312 (1990).
9. Y. Saito and S. Yuta, Teaching for mobile robot-autonomous navigation based on taught path and environment by radio control, in: *Proc. 9th Annu. Conf. of the Robotics Society of Japan*, pp. 255–258 (1991).
10. I. Ohya, A. Kosaka, and A. Kak, Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing, *IEEE Trans. Robotics Automat.* **14**, 969–978 (1998).
11. A. Kotanen, M. Hannikainen, H. Leppakoski and T. Hamalainen, Positioning with IEEE 802.11b wireless LAN, in: *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications*, Beijing, pp. 2218–2222 (2003).
12. A. M. Ladd, K. E. Bekris, A. P. Rudys, D. S. Wallach, and L. E. Kavraki, On the feasibility of using wireless Ethernet for indoor localization, *IEEE Trans. Robotics Automat.* **20**, 555–559 (2004).
13. D. Haehnel, W. Burgard, D. Fox, K. P. Fishkin and M. Philipose, Mapping and localization with RFID technology, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, New Orleans, LA, pp. 1015–1020 (2004).
14. D. C. Q. Chen and V. Thomas, Optimization of inductive RFID technology, in: *Proc. IEEE Int. Symp. on Electronics and Environment*, Denver, CO, pp. 82–87 (2001).
15. N. Winters, J. Gaspar, G. Lacey and J. S. Victor, Omni-directional vision for robot navigation, in *Proc. IEEE Workshop Omnidirectional Vision*, Hilton Head Island, SC, pp. 21–28 (2000).
16. Y. Matsumoto, K. Ikeda, M. Inaba and H. Inoue, Visual navigation using omnidirectional view sequence, in: *Proc. IEEE Int. Conf. on Intelligent Robots and Systems*, Kyongju, pp. 317–322 (1999).
17. K. Martin, Estimating azimuth and elevation from interaural difference, in: *Proc. IEEE Workshop on Applications of Signal Processing to Acoustics and Audio*, New Paltz, pp. 15–18 (1995).
18. C. J. MacCabe and D. Furlong, Virtual imaging capabilities of surround sound systems, *J. Audio Eng. Soc.* **42**, 38–49 (1994).
19. K. Nakadai, H. G. Okuno and H. Kitano, Epipolar geometry based sound localization and extraction for humanoid audition, in: *Proc. Intelligent Robots and Systems*, Hawaii, HI, vol. 2, pp. 1395–1401 (2001).

20. H. G. Okuno, K. Nakadai, K. I. Hidai, H. Mizoguchi and H. Kitano, Human–robot interaction through real-time auditory and visual multiple-talker tracking, in: *Proc. Intelligent Robots and Systems*, Hawaii, HI, vol. 2, pp. 1402–1409 (2001).
21. K. Nakadai, D. Matsuura, H. G. Okuno, and H. Tsujino, Improvement of recognition of simultaneous speech signals using AV integration and scattering theory for humanoid robots, *Speech Commun.* **44**, 97–112 (2004).
22. G. Xuan, W. Zhang and P. Chai, EM algorithms of Gaussian mixture model and hidden Markov model, in: *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, pp. 145–148 (2001).
23. M. Wax and T. Kailath, Optimal localization of multiple sources by passive arrays, *IEEE Trans. Acoust., Speech Signal Process.* **31**, 1210–1217 (1983).
24. S. Kagami, Y. Tamai, H. Mizoguchi and T. Kanade, Microphone array for 2D sound localization and capture, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, Barcelona, pp. 703–708 (2004).
25. C. H. Knapp and G. C. Carter, The generalized correlation method for estimation of time delay, *IEEE Trans. Acoust., Speech Signal Process.* **24**, 320–327 (1976).
26. G. C. Carter, A. H. Nuttall and P. G. Cable, The smoothed coherence transform, *IEEE Signal Process. Lett.* **61**, 1497–1498 (1973).
27. M. Wax, T. J. Shan, and T. Kailath., Spatio-temporal spectral analysis by eigenstructure methods, *IEEE Trans. Acoust. Speech Signal Process.* **32**, 817–827 (1984).
28. M. S. Brandstein and H. F. Silverman, A robust method for speech signal time-delay estimation in reverberant rooms, in: *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Munich, pp. 375–378 (1997).
29. M. Ricci, G. M. Edelman and J. Wray, Adaptation of orienting behavior: from the Barn Owl to a robotic system, *IEEE Trans. Robotics Automat.* **15**, 96–110 (1999).
30. S. B. Andersson, A. A. Handzel, V. Shah and P. S. Krishnaprasad, Robot phonotaxis with dynamic sound-source localization, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, Barcelona, pp. 4833–4838 (2004).
31. S. S. Ge, A. P. Loh, and F. Guan, Sound localization based on mask diffraction, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, Taipei, pp. 1972–1977 (2003).
32. A. D. Horchler, R. E. Reeve, B. Webb, and R. D. Quinn, Robot phonotaxis in the wild: a biological inspired approach to outdoor sound localization, *Adv. Robotics* **18** 801-816 (2004).
33. E. Mumlol, M. Nolich and G. Vercelli, Algorithms for acoustic localization based on microphone array in service robotics, *Robotics Autonomous Syst.* **42**, 69–88 (2003).
34. Q. H. Wang, T. Ivanov and P. Aarabi, Acoustic robot navigation using distributed microphone arrays, *Information Fusion* **5**, 131–140 (2004).
35. J.-M. Valin, F. Michaud, B. Hadjou and J. Rouat, Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, Barcelona, pp. 1033–1038 (2004).
36. B. C. Ng and C. M. S. See, Sensor-array calibration using a maximum-likelihood approach, *IEEE Trans. Antennas Propag.* **44**, 827–835, (1996).
37. D. B. Ward, E. A. Lehmann and R. C. Williamson, Particle filtering algorithms for tracking an acoustic source in a reverberant environment, *IEEE Trans. Speech Audio Process.* **11**, 826–836 (2003).
38. B. H. Lee, Embedded internet systems: poised for takeoff, *IEEE Internet Comput.* **2**, 24–29 (1998).
39. J. Bentham, *TCP/IP Lean: Web Servers for Embedded Systems*. CMP Books, Lawrence, KS (2000).
40. L.-W. Wu and J.-S. Hu, Distributed embedded real-time Ethernet platform for robots control, in: *Proc. IEEE Int. Conf. ICM/HIMA*, Taipei, pp. 370–375 (2005).
41. D. Golubovic, B. Li and H. Hu, A hybrid software platform for Sony AIBO robots, in: *Proc. 7th Int. Symp. on RobotCup*, Padua, pp. 478–486 (2003).

42. D. A. Reynolds and R. C. Rose, Robust text-independent speaker independent using Gaussian mixture speaker models, *IEEE Trans. Speech Audio Process.* **3**, 72–83 (1995).
43. J. B. MacQueen, Some methods for classification and analysis of multivariate observations, in: *Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability*, Berkeley, CA, pp. 281–297 (1967).
44. C. Elkan, Using the triangle inequality to accelerate *k*-means, in: *Proc. 20th Int. Conf. on Machine Learning*, Washington, DC, pp. 147–153 (2003).
45. M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, New York, NY (2001).
46. F. Asano, Y. Motomura, H. Asoh, T. Yoshimura, N. Ichimura, K. Yamamoto, N. Kitawaki and S. Nakamura, Detection and separation of speech segment using audio and video information fusion, in: *Proc. Eurospeech*, Geneva, pp. 2257–2260 (2003).
47. L.-W. Wu and J.-S. Hu, Robot control system using embedded network, *Taiwan Patent I238622* (2004).

APPENDIX: PROOF OF (22) AND (23)

The problem is formulated as:

$$\min\left\{\sum_{q_\mathrm{p}} \alpha_\mathrm{p}\{\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})\mathbf{U}\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})^\mathrm{T}\} + \sum_{q_\mathrm{M}} \alpha_\mathrm{M}\{\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})\mathbf{U}\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})^\mathrm{T}\}\right\}$$
$$\text{s.t.}\quad \alpha_\mathrm{P}\alpha_\mathrm{M} = 1, \alpha_\mathrm{P} > 0, \alpha_\mathrm{M} > 0.$$

According to the constraint, set $\alpha_\mathrm{M} = 1/\alpha_\mathrm{P}$. Then, the cost function becomes:

$$\min\left\{\sum_{q_\mathrm{p}} \alpha_\mathrm{p}\{\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})\mathbf{U}\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})^\mathrm{T}\} + \sum_{q_\mathrm{M}} \frac{1}{\alpha_\mathrm{p}}\{\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})\mathbf{U}\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})^\mathrm{T}\}\right\}.$$

Setting the first derivative with respect to α_P be zero gives:

$$\sum_{q_\mathrm{P}} \mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})\mathbf{U}\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})^\mathrm{T} - \sum_{q_\mathrm{M}} \alpha_\mathrm{p}^{-2}\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})\mathbf{U}\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})^\mathrm{T} = 0.$$

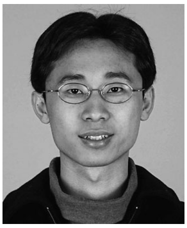
Therefore:

$$\alpha_\mathrm{P} = \sqrt{\frac{\sum_{q_\mathrm{M}} \mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})\mathbf{U}\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})^\mathrm{T}}{\sum_{q_\mathrm{p}} \mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})\mathbf{U}\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})^\mathrm{T}}}$$
$$\alpha_\mathrm{M} = \sqrt{\frac{\sum_{q_\mathrm{p}} \mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})\mathbf{U}\mathbf{C}_\mathrm{P}(\boldsymbol{q}_\mathrm{P})^\mathrm{T}}{\sum_{q_\mathrm{M}} \mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})\mathbf{U}\mathbf{C}_\mathrm{M}(\boldsymbol{q}_\mathrm{M})^\mathrm{T}}}.$$

ABOUT THE AUTHORS

Li-Wei Wu was born in Chiayi, Taiwan, in 1974. He received his BS degree in Mechanical Engineering from Chinese Culture University, Taiwan, ROC, in 1998, and his MS degree in Mechanical and Aerospace Engineering from Chung Hua University, Taiwan, ROC, in 2000. He received his Ph.D. from the Department of Electrical and Control Engineering at National Chiao-Tung University in 2006. He received the Honorable Mention of the Chinese Taiwan Ministry of Education (championship) in the 2003 Embedded Software Design Contest, sponsored by the Ministry of Education Advisory Office and the System-on-a-Chip Consortium. He also received the Honorable Mention of the Chinese Taiwan Ministry of Education in the 2005 Intelligent Robot Contest. His research interests include robotics, real-time network protocol, embedded Ethernet, real-time systems, and embedded system design and implementation. Currently he is a Senior Engineer at Compal Communications.

Wei-Han Liu was born in Kaohsiung, Taiwan, in 1977. He received his BS and MS degrees in Electrical and Control Engineering from National Chiao Tung University, Taiwan, ROC, in 2000 and 2002. He is currently a PhD candidate in the Department of Electrical and Control Engineering at the National Chiao Tung University, Taiwan, ROC. He is the Champion of the TI DSP Solutions Design Challenge in 2000 and of the national competition held by the Ministry of Education Advisor Office in 2001. He is the winner of the best paper award at IEEE/ASME 2002. His research interests include sound source localization, microphone array signal processing, adaptive signal processing, speech signal processing and robot localization.



Chieh-Cheng Cheng was born in 1978. He received his BS and PhD degrees in Electrical and Control Engineering from the National Chiao Tung University, Taiwan, ROC, in 2000 and 2006. He is the champion of the TI DSP Solutions Design Challenge in 2000 and of the national competition held by the Ministry of Education Advisor Office in 2001. His research interests include sound source localization, microphone array signal processing, adaptive signal processing, pattern recognition, speech signal processing, and echo and noise cancellation. Currently he is a Senior Engineer at MediaTek.

Jwu-Sheng Hu was born in Taipei, Taiwan in 1962. He received his BS degree from the Department of Mechanical Engineering, National Taiwan University, Taiwan, in 1984, and his MS and PhD degrees from the Department of Mechanical Engineering, University of California at Berkeley, in 1988 and 1990, respectively. He is currently a Professor in the Department of Electrical and Control Engineering, National Chiao-Tung University, Taiwan, ROC. His current research interests include microphone array signal processing, active noise control, embedded system design and robotics.