Available online at www.sciencedirect.com

ScienceDirect

Pattern Recognition Letters 29 (2008) 149–160

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Indoor sound field feature matching for robot's location and orientation detection

Jwu-Sheng Hu [1], Wei-Han Liu *, Chieh-Cheng Cheng [2]

Department of Electrical and Control Engineering, National Chiao-Tung University, 1001 Ta Hsueh Road, Hsinchu 300, Taiwan, ROC

Received 7 December 2006; received in revised form 16 July 2007
Available online 25 September 2007

Communicated by O. Siohan

Abstract

In this work, an indoor sound field feature matching method is proposed and is applied to detect a mobile robot's location and orientation. The sound field feature, captured from a sound source to a pair of microphones, contains the dynamic of the propagation path. Because of the complexity of indoor environment, the features from different path can be distinguished using appropriate models. Gaussian mixture models are utilized in this paper to characterize the phase difference and magnitude ratio distributions between the microphone pair in consecutive data frames. The application provides an alternative thinking compared with traditional methods such as direction of arrival (DOA) using propagation delay. They usually suffer from reverberation, non-line-of-sight and microphone mismatch problems. The experimental results show the method not only has a high recognition rate for robot's location and orientation, but also is robust against environmental noise.

Keywords: GMM; Robot localization; Robot's orientation detection

1. Introduction

Indoor robot localization is an important issue in the field of robotics. Various equipments, such as camera, radio frequency identification (RFID), infrared red (IR), ultra sonic sensor, laser, wireless LAN-based methods and inertial navigation sensor have been adopted to provide different solutions (Borenstein et al., 1996; Georgiev and Allen, 2004; Gutierrez-Osuna et al., 1998; Ladd et al., 2004; Larsson et al., 1996; Lee et al., 2003; McGillem and Rappaport, 1988; Ohya et al., 1998). Pattern matching or pattern recognition-based algorithms are also proposed in this research domain. Vlassis et al. (2001) utilized edge-based feature vectors of the omni-directional images for robot localization. A place recognition method based on image signature matching was presented for mobile robots (Argamon-Engelson, 1998). For range-finder-based sensors, Weiss et al. (1994) proposed a method based on matching two scan results to derive the position and orientation of a moving indoor system.

For indoor robots, audio devices such as loudspeakers and microphones are becoming basic equipments. These sound-related devices can generally provide a more nature way for robots to communicate with human. Additionally, some researchers believe that these devices can be utilized for robot localization (Tamai et al., 2004a; Wang et al., 2004). This work investigates the feasibility of using sound field feature matching for robot's location and orientation detection and proposes a robust sound-based indoor robot's pose detection system utilizing two microphones.

* Corresponding author. Tel.: +886 3 5712121 54424; fax: +886 3 5715998.

E-mail addresses: jshu@cn.nctu.edu.tw (J.-S. Hu), lukeliu.ece89g@nctu.edu.tw (W.-H. Liu), canson.ece89g@nctu.edu.tw (C.-C. Cheng).

[1] Tel.: +886 3 5712121 54318; fax: +886 3 5715998.

[2] Tel.: +886 3 5712121 54424; fax: +886 3 5715998.

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2007.09.007

1.1. Traditional sound-based robot localization methods and known problems

The idea of using multiple microphones to localize sound sources has been developed for a long time. Among various kinds of sound source localization methods, generalized cross-correlation (GCC)-based methods (Brandstein and Silverman, 1997; Carter et al., 1973; Knapp and Carter, 1976; Nikas and Shao, 1995) were discussed for robot localization application (Wang et al., 2004). In general, sound-based robot localization system uses a speaker mounted on the robot to produce sound and estimates the location of the sound source, which is the robot's location, by a set of microphone array installed in the room (Tamai et al., 2004a; Wang et al., 2004). The main difficulty for indoor robot localization using sound wave is the complex propagation behavior such as reflection and diffraction. Theoretically, the values of phase difference and magnitude ratio among microphones are directly related to the sound wave arrival direction and the distance between a sound source and microphones. However, these straightforward relations only exist in free space or environments with simple geometry. In real environments, these values exhibit stochastic phenomena due to the distributed nature of the propagation path dynamics and the limitation of finite-length data. Furthermore, complex boundary conditions, near-field effect, and local sound scattering make these values hard to correlate with the source location. These variations generally result in uncertain estimation errors and make sound-based localization methods unreliable. Moreover, for indoor applications, the robot may move to a location that is non-line-of-sight to the sensors, i.e., without direct paths between the robot and microphones. Under this circumstance, traditional methods cannot locate the robot accurately.

Another well-known problem of sound-based robot localization methods is the microphone mismatch problem. If the microphones are not mutually matched, then the phase difference information among microphones may be distorted. However, pre-matched microphones are relatively expensive and mismatched microphones are difficult to calibrate accurately since the characteristics of microphones change with the sound directions. Consequently, the estimation accuracy varies from different microphone pairs and is difficult to be evaluated.

1.2. The proposed method based on sound field feature matching

Traditional sound source localization algorithms attempt to suppress the effects of complex propagation behavior, as well as estimate the direction of the direct sound source. Unlike existing sound-based robot localization systems which focus on eliminating the influence of reflection and diffraction, this work treats the propagation behavior as a local feature and attempts to recognize it by pattern matching method. In practice, the complex propagation behavior of a sound sources results in location or orientation dependent phase difference and magnitude ratio distributions. For example, Figs. 1 and 2 show the histograms of phase difference and the magnitude ratio in consecutive data frames measured between a microphone pair for the location "A" in a line-of-sight case and the location "B" in a non-line-of-sight case (the figure of the experimental environment is shown in Fig. 9). Obviously, even under the line-of-sight case, the values of phase difference and magnitude ratio are not fixed due to the complex propagation behavior.

The examples of sound field features given in Figs. 1 and 2 are used to mark the location or orientation of a sound source that is mounted on the robot. Notably, both the magnitude ratio and the phase difference between two microphones are content independent. In other words, the content of the sound produced by the robot does not have to be defined. For example, the sound can be conversation, or even the noise emitted by an autonomous vacuum-cleaning robot. This work adopts Gaussian mixture models (GMMs) (Reynolds and Rose, 1995) to model phase difference and magnitude ratio distributions and proposes two models, robot localization model (RLM) and robot orientation model (ROM). The first model (RLM) is used for robot's location detection and the second model (ROM) is used for robot's orientation detection. The unique advantage of the proposed method is the detection of location and orientation in non-line-of-sight cases, i.e., when no direct path is available between the robot and

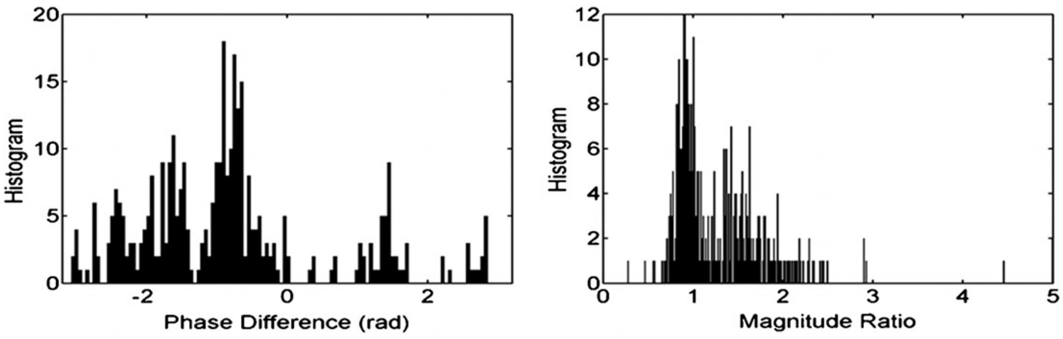

Fig. 1. Phase difference and magnitude ratio in line-of-sight case.

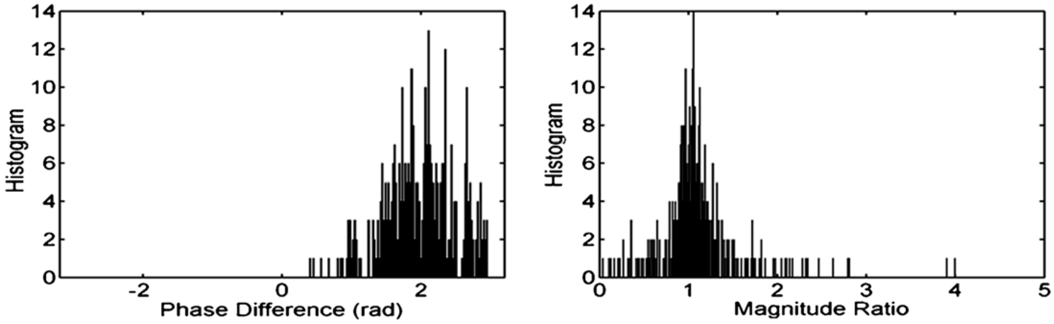

Fig. 2. Phase difference and magnitude ratio in non-line-of-sight case.

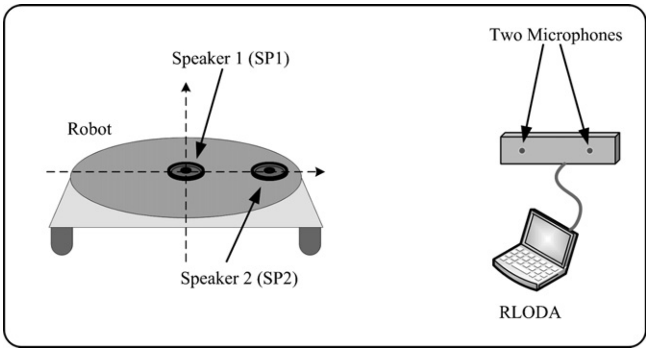

Fig. 3. Speaker and microphone configuration of the proposed system.

the microphones. To adapt to the environmental noises and enhance the robustness of the feature identification, an on-line calibration procedure is also proposed.

The remainder of this paper is organized as follows. The next section introduces the overall system architecture. Section 3 describes the design of the directional sound pattern for orientation detection. Section 4 presents the formulations of the proposed RLM and ROM. The experimental results are discussed in Section 5 and, finally, conclusions are drawn in Section 6.

2. System architecture

As shown in Fig. 3, the proposed system contains two speakers on the robot and a robot's location and orientation detection agent (RLODA) with two microphones. The RLODA can be placed in any part of the room as long as the reception of sound from the robot is clear enough.[3] The sound pattern generated by Speaker 1 (SP1) is received by the RLODA and the RLMs for different sound source locations can be obtained by modeling the location dependent sound field features (phase difference and magnitude ratio distributions) measured between the two microphones. When the system attempts to build the ROMs, both SP1 and SP2 are used to generate a directional sound pattern. Note that the detail of generating a directional sound pattern is described in Section 3. Because the sound pattern generated by SP1 and SP2 is directional, the sound field features change with the robot's orientation and can be utilized for orientation detection.

Fig. 4 depicts the overall system architecture. Stage I in Fig. 4 is the pre-recording stage, in which the robot moves and changes its orientation in the environment when the environment is quiet, and produces sound through the speakers to obtain a pre-recorded database. Since the sound is recorded by the two microphones, the information of the sound field features can be obtained by this database.

Once the pre-recording stage is finished, the system enters Stage II called silent stage. In this stage, the robot remains silent and the RLODA records the environmental noises. Assuming that noise signals are additive, the sound recorded in real application can be considered as the linear combination of robot's sound and environmental noises. Therefore, this stage adds the environmental noises to the pre-recorded database to construct the training features, phase difference and magnitude ratio distributions, and then utilizes these features to trains the parameters of RLMs and ROMs. Through this process, the effect of environmental noises is adapted in this stage.

When the robot needs to know its location or orientation, the system then switches to the sounding stage, in which the robot produces a sound into the room for the RLODA to detect the robot's location or orientation. If the robot's location is required, the SP1 is used to generate sound; conversely, both SP1 and SP2 are excited if the robot's orientation is needed. Because the microphones used in these three stages are the same, the mismatched characteristics between microphones are collected in the pre-recording database and would not influence the detection results of proposed system. The sounding and the silent stages can be switched to each other iteratively for location or orientation detection and environmental noises adaptation. Fig. 5 illustrates the flowchart of proposed system.

Additionally, wireless communication technologies such as Wireless Ethernet can be adopted to accomplish the stage synchronization and communication between the robot and the RLODA.

[3] In this paper, we do not discuss the issue of placement of RLODA.

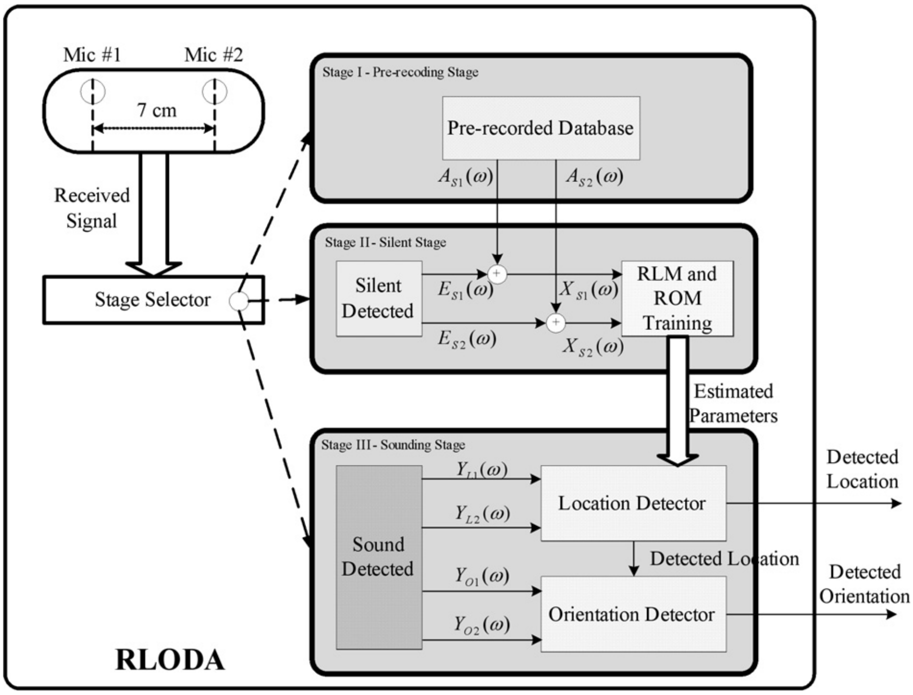

Fig. 4. Overall system architecture.

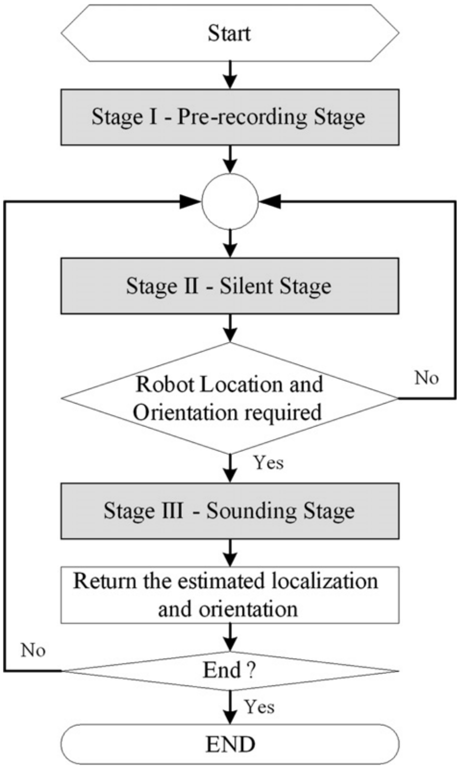
Start
Stage I - Pre-recording Stage
Stage II - Silent Stage
Robot Location and Orientation required
No
Yes
Stage III - Sounding Stage
Return the estimated localization and orientation
No
End ?
Yes
END

Fig. 5. Flowchart of the proposed system.

3. Directional sound pattern design for robot orientation detection

To detect the robot's orientation by the sound field features, the sound pattern generated by the robot should be correlated with the robot's orientation. However, a general omni-directional sound pattern may lead to the same sound fields when the robot changes its orientation because the emitted sound has the same characteristics in all directions. Therefore, a directional sound emission approach must be designed. To realize a directional sound pattern, the idea of speaker array beamforming (Tamai et al., 2004b; Yamada et al., 2004) is adopted in this work to guarantee the directivity of the generated sound pattern. Besides directivity, another constraint on the generated sound pattern is the number of symmetric axes (β) in the horizontal plane. Fig. 6 shows an example of how β affects the orientation detection, where the solid line denotes the generated sound pattern, the dotted line denotes the symmetric axes, and the arrow denotes the robot's orientation.

As shown in Fig. 6, the sound patterns generated when the robot's orientation is 0°, 90°, 180°, and 270° are exactly the same when $\beta = 4$. A sound pattern generated when the robot points at a certain direction (0° in the example) would have $\beta - 1$ identical sound patterns. Therefore, the generated sound could only be symmetrical along one axis ($\beta = 1$) to avoid confusion in orientation detection. Consequently, this work proposes a method that utilizes two speakers to generate the sound pattern that conforms to the constraint by

$$\begin{aligned} J_{\mathrm{SP1}}(n) &= J(n) \\ J_{\mathrm{SP2}}(n) &= 0.5 \times J(n) \end{aligned} \tag{1}$$

where $J(n)$ is the original sound source and $J_{\mathrm{SP1}}(n)$ and $J_{\mathrm{SP2}}(n)$ is the sound emitted by SP1 and SP2. The distance between two speakers is set to 0.2 m.

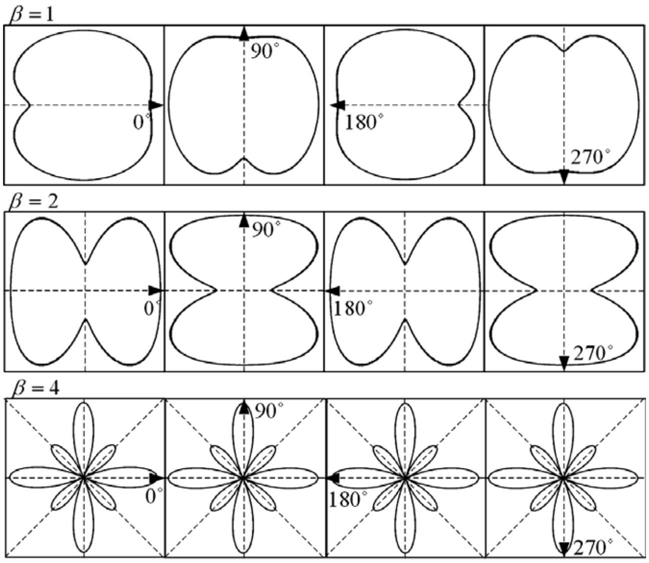

Fig. 6. Relations between β and the sound pattern.

Fig. 7 depicts the simulation of the generated sound pattern of the proposed system based on the sound propagation theories introduced by Parker (1988) when the robot's orientation is 0°, where the sound power is measured at 1 m away from the SP1 with the same height. The solid lines in the circle depict the relative sound power in dB. As shown in Fig. 7, the generated sound pattern is symmetric along only one axis and is suitable for robot's orientation detection.

4. Robot localization model (RLM) and robot orientation model (ROM)

4.1. A description of the proposed RLM and ROM

To establish both RLMs and ROMs, the RLODA needs to construct models for the sound fields at different locations and orientations. $P_{\mathrm{Sx}}(\omega_b)$ and $M_{\mathrm{Sx}}(\omega_b)$ denote the phase difference and magnitude ratio, respectively, for constructing RLM ($S = L$) or ROM ($S = O$) at frequency ω_b, $b \in \{1, \ldots, B\}$. The GMMs are defined as the weighted sum of N_1 and N_2 mixtures of Gaussian component densities shown below,

$$G(\boldsymbol{P}_{\mathrm{Sx}}|\boldsymbol{\lambda}_{\mathrm{SP}}) = \sum_{i=1}^{N_1} \rho_{\mathrm{SP},i} g_i(\boldsymbol{P}_{\mathrm{Sx}}) \tag{2}$$

$$G(\boldsymbol{M}_{\mathrm{Sx}}|\boldsymbol{\lambda}_{\mathrm{SM}}) = \sum_{i=1}^{N_2} \rho_{\mathrm{SM},i} g_i(\boldsymbol{M}_{\mathrm{Sx}}) \tag{3}$$

where $S = \{L, O\}$, $\boldsymbol{P}_{\mathrm{Sx}} = [P_{\mathrm{Sx}}(\omega_1) \quad \cdots \quad P_{\mathrm{Sx}}(\omega_B)]^{\mathrm{T}}$, $\boldsymbol{M}_{\mathrm{Sx}} = [M_{\mathrm{Sx}}(\omega_1) \quad \cdots \quad M_{\mathrm{Sx}}(\omega_B)]^{\mathrm{T}}$. $\rho_{\mathrm{SP},i}$ and $\rho_{\mathrm{SM},i}$ are the ith mixture weights, and $g_i(\boldsymbol{P}_{\mathrm{Sx}})$ and $g_i(\boldsymbol{M}_{\mathrm{Sx}})$ are the Gaussian density function. Notably, the mixture weights must satisfy the constraints:

$$\sum_{i=1}^{N_1} \rho_{\mathrm{SP},i} = 1 \quad \text{and} \quad \sum_{i=1}^{N_2} \rho_{\mathrm{SM},i} = 1 \tag{4}$$

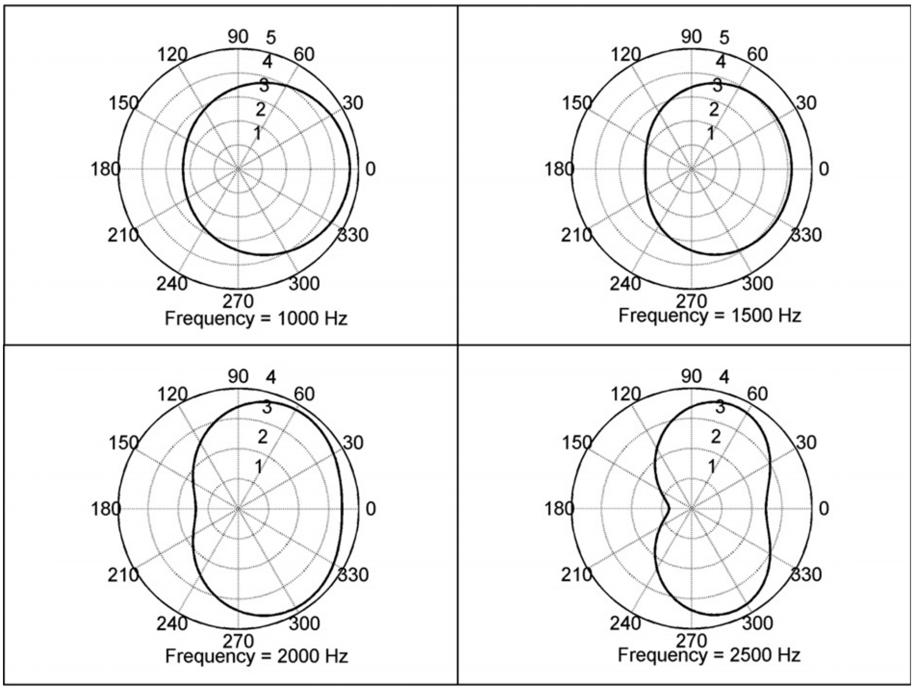

Fig. 7. Simulation of generated sound pattern.

The terms $\boldsymbol{\lambda}_{\mathrm{SP}}$ and $\boldsymbol{\lambda}_{\mathrm{SM}}$ represent the parameters of N_1 and N_2 component densities:

$$\boldsymbol{\lambda}_{\mathrm{SP}} = \{\boldsymbol{\rho}_{\mathrm{SP}}, \boldsymbol{\mu}_{\mathrm{SP}}, \boldsymbol{\Sigma}_{\mathrm{SP}}\} \quad \text{and} \quad \boldsymbol{\lambda}_{\mathrm{SM}} = \{\boldsymbol{\rho}_{\mathrm{SM}}, \boldsymbol{\mu}_{\mathrm{SM}}, \boldsymbol{\Sigma}_{\mathrm{SM}}\} \tag{5}$$

where

$\boldsymbol{\rho}_{\mathrm{SP}} = \lfloor \rho_{\mathrm{SP},1} \quad \cdots \quad \rho_{\mathrm{SP},N_1} \rfloor$ denotes the phase difference mixture weight vector with dimensions $1 \times N_1$.
$\boldsymbol{\rho}_{\mathrm{SM}} = \lfloor \rho_{\mathrm{SM},1} \quad \cdots \quad \rho_{\mathrm{SM},N_2} \rfloor$ denotes the magnitude ratio mixture weight vector with dimensions $1 \times N_2$.
$\boldsymbol{\mu}_{\mathrm{SP}} = \lfloor \boldsymbol{\mu}_{\mathrm{SP},1} \quad \cdots \quad \boldsymbol{\mu}_{\mathrm{SP},N_1} \rfloor$ denotes the phase difference mean matrix with dimensions $B \times N_1$.
$\boldsymbol{\mu}_{\mathrm{SM}} = \lfloor \boldsymbol{\mu}_{\mathrm{SM},1} \quad \cdots \quad \boldsymbol{\mu}_{\mathrm{SM},N_2} \rfloor$ denotes the magnitude ratio mean matrix with dimensions $B \times N_2$.
$\boldsymbol{\Sigma}_{\mathrm{SP}} = \lfloor \boldsymbol{\Sigma}_{\mathrm{SP},1} \quad \cdots \quad \boldsymbol{\Sigma}_{\mathrm{SP},N_1} \rfloor$ denotes the phase difference covariance matrix with dimensions $B \times BN_1$.
$\boldsymbol{\Sigma}_{\mathrm{SM}} = \lfloor \boldsymbol{\Sigma}_{\mathrm{SM},1} \quad \cdots \quad \boldsymbol{\Sigma}_{\mathrm{SM},N_2} \rfloor$ denotes the magnitude ratio covariance matrix with dimensions $B \times BN_2$.

The parameters $\boldsymbol{\lambda}_{\mathrm{SP}}$ and $\boldsymbol{\lambda}_{\mathrm{SM}}$ in (5) can be estimated by the iterative EM algorithm (Xuan et al., 2001) which guarantees a monotonic increase in the model's log-likelihood value. The iterative procedure can be divided into the following two steps:

Expectation step:

$$G\left(i|\boldsymbol{P}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SP}}\right) = \boldsymbol{\rho}_{\mathrm{SP},i} g_i\left(\boldsymbol{P}_{\mathrm{Sx}}^{(t)}\right) \Big/ \sum_{i=1}^{N_1} \boldsymbol{\rho}_{\mathrm{SP},i} g_i\left(\boldsymbol{P}_{\mathrm{Sx}}^{(t)}\right) \tag{6}$$

$$G\left(i|\boldsymbol{M}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SM}}\right) = \boldsymbol{\rho}_{\mathrm{SM},i} g_i\left(\boldsymbol{M}_{\mathrm{Sx}}^{(t)}\right) \Big/ \sum_{i=1}^{N_2} \boldsymbol{\rho}_{\mathrm{SM},i} g_i\left(\boldsymbol{M}_{\mathrm{Sx}}^{(t)}\right) \tag{7}$$

where $G\left(i|\boldsymbol{P}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SP}}\right)$ and $G\left(i|\boldsymbol{M}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SM}}\right)$ are posteriori probabilities.

Maximization step:

(i) Estimate the mixture weights:

$$\rho_{\mathrm{SP},i} = 1 \Big/ T \sum_{t=1}^{T} G\left(i|\boldsymbol{P}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SP}}\right) \tag{8}$$

$$\rho_{\mathrm{SM},i} = 1 \Big/ T \sum_{t=1}^{T} G\left(i|\boldsymbol{M}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SM}}\right) \tag{9}$$

(ii) Estimate the mean vector:

$$\boldsymbol{\mu}_{\mathrm{SP},i} = \sum_{t=1}^{T} G\left(i|\boldsymbol{P}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SP}}\right) \boldsymbol{P}_{\mathrm{Sx}}^{(t)} \Big/ \sum_{t=1}^{T} G\left(i|\boldsymbol{P}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SP}}\right) \tag{10}$$

$$\boldsymbol{\mu}_{\mathrm{SM},i} = \sum_{t=1}^{T} G\left(i|\boldsymbol{M}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SM}}\right) \boldsymbol{M}_{\mathrm{Sx}}^{(t)} \Big/ \sum_{t=1}^{T} G\left(i|\boldsymbol{M}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SM}}\right) \tag{11}$$

(iii) Estimate the variances:

$$\sigma_{\mathrm{SP},i}^{2}(\omega_b) = \left(\sum_{t=1}^{T} G\left(i|\boldsymbol{P}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SP}}\right) P_{\mathrm{Sx}}^{(t)2}(\omega_b) \Big/ \sum_{t=1}^{T} G\left(i|\boldsymbol{P}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SP}}\right)\right) - \mu_{\mathrm{SP},i}^{2}(\omega_b) \tag{12}$$

$$\sigma_{\mathrm{SM},i}^{2}(\omega_b) = \left(\sum_{t=1}^{T} G\left(i|\boldsymbol{M}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SM}}\right) M_{\mathrm{Sx}}^{(t)2}(\omega_b) \Big/ \sum_{t=1}^{T} G\left(i|\boldsymbol{M}_{\mathrm{Sx}}^{(t)}, \boldsymbol{\lambda}_{\mathrm{SM}}\right)\right) - \mu_{\mathrm{SM},i}^{2}(\omega_b) \tag{13}$$

However, the EM algorithm is sensitive to the choice of initial model. A good choice of initial model results in a lower number of iterations of the EM algorithm. *K*-means related approaches are known to be effective in finding a suitable initial model (MacQueen, 1967). This work utilizes an accelerated *K*-means algorithm proposed by Elkan (2003), which can significantly reduce the computational power requirement.

The proposed RLM and ROM at each location and orientation are defined as the linear combination of the phase difference GMM and the magnitude ratio GMM:

$$F_{\mathrm{RLM}} = \alpha_{\mathrm{LP}} G(\boldsymbol{P}_{\mathrm{Lx}}|\boldsymbol{\lambda}_{\mathrm{LP}}) + \alpha_{\mathrm{LM}} G(\boldsymbol{M}_{\mathrm{Lx}}|\boldsymbol{\lambda}_{\mathrm{LM}}) \tag{14}$$

$$F_{\mathrm{ROM}} = \alpha_{\mathrm{OP}} G(\boldsymbol{P}_{\mathrm{Ox}}|\boldsymbol{\lambda}_{\mathrm{OP}}) + \alpha_{\mathrm{OM}} G(\boldsymbol{M}_{\mathrm{Ox}}|\boldsymbol{\lambda}_{\mathrm{OM}}) \tag{15}$$

where α_{LP}, α_{OP}, α_{LM} and α_{OM} represent the weighting factors. The values of α_{SP} and α_{SM} can be chosen based on the sum of the correlation values among trained locations of the phase difference GMM and magnitude ratio GMM. The GMM with higher correlation summation would be assigned a lower weight, since the ability to discriminate is considered lower under this circumstance, and vice versa. Under this principle, α_{SP} and α_{SM} are determined by the following formula:

$$\min\left\{\sum_{\boldsymbol{q}_{\mathrm{SP}}} \alpha_{\mathrm{SP}}\{\mathbf{C}_{\mathrm{SP}}(\boldsymbol{q}_{\mathrm{SP}})\mathbf{U}\mathbf{C}_{\mathrm{SP}}(\boldsymbol{q}_{\mathrm{SP}})^{\mathrm{T}}\} + \sum_{\boldsymbol{q}_{\mathrm{SM}}} \alpha_{\mathrm{SM}}\{\mathbf{C}_{\mathrm{SM}}(\boldsymbol{q}_{\mathrm{SM}})\mathbf{U}\mathbf{C}_{\mathrm{SM}}(\boldsymbol{q}_{\mathrm{SM}})^{\mathrm{T}}\}\right\} \tag{16}$$

$$\text{s.t.} \quad \alpha_{\mathrm{SP}}\alpha_{\mathrm{SM}} = 1, \quad \alpha_{\mathrm{SP}} > 0, \quad \alpha_{\mathrm{SM}} > 0$$

where $\boldsymbol{q}_{\mathrm{SP}} \in Q_{\mathrm{SP}}$ and $\boldsymbol{q}_{\mathrm{SM}} \in Q_{\mathrm{SM}}$ are the B dimensional random vectors in the operation ranges, Q_{SP} and Q_{SM}.

$$\mathbf{C}_{\mathrm{SP}}(\boldsymbol{q}_{\mathrm{SP}}) = [C(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(1)) \quad C(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(2)) \quad \cdots \quad C(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(L))],$$

$$\mathbf{C}_{\mathrm{SM}}(\boldsymbol{q}_{\mathrm{SM}}) = [C(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(1)) \quad C(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(2)) \quad \cdots \quad C(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(L))],$$

and

$$\mathbf{U}=\begin{bmatrix}0&1&1&\cdots&\cdots&1\\0&0&1&1&\cdots&1\\\vdots&0&0&\ddots&\cdots&1\\\vdots&\vdots&0&\ddots&1&1\\\vdots&\vdots&\vdots&\vdots&0&1\\0&0&0&0&0&0\end{bmatrix}$$

with dimension $L\times L$.

In addition,

$$C(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(l))=H(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(l))\Bigg/\sqrt{\sum_{\boldsymbol{q}_{\mathrm{SP}}}H^{2}(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(l))},\tag{17}$$

$$C(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(l))=H(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(l))\Bigg/\sqrt{\sum_{\boldsymbol{q}_{\mathrm{SM}}}H^{2}(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(l))},\tag{18}$$

$$H(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(l))=G(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(l))-\left(\sum_{\boldsymbol{q}_{\mathrm{SP}}}G(\boldsymbol{q}_{\mathrm{SP}}|\boldsymbol{\lambda}_{\mathrm{SP}}(l))/N(\boldsymbol{q}_{\mathrm{SP}})\right),$$

and

$$H(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(l))=G(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(l))-\left(\sum_{\boldsymbol{q}_{\mathrm{SM}}}G(\boldsymbol{q}_{\mathrm{SM}}|\boldsymbol{\lambda}_{\mathrm{SM}}(l))/N(\boldsymbol{q}_{\mathrm{SM}})\right)\tag{19}$$

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

The values of α_{SP} and α_{SM} can be obtained by solving (16) as:

$$\alpha_{\mathrm{SP}}=\sqrt{\sum_{\boldsymbol{q}_{\mathrm{SM}}}\mathbf{C}_{\mathrm{SM}}(\boldsymbol{q}_{\mathrm{SM}})\mathbf{U}\mathbf{C}_{\mathrm{SM}}(\boldsymbol{q}_{\mathrm{SM}})^{\mathrm{T}}\Bigg/\sum_{\boldsymbol{q}_{\mathrm{SP}}}\mathbf{C}_{\mathrm{SP}}(\boldsymbol{q}_{\mathrm{SP}})\mathbf{U}\mathbf{C}_{\mathrm{SP}}(\boldsymbol{q}_{\mathrm{SP}})^{\mathrm{T}}}\tag{20}$$

$$\alpha_{\mathrm{SM}}=\sqrt{\sum_{\boldsymbol{q}_{\mathrm{SP}}}\mathbf{C}_{\mathrm{SP}}(\boldsymbol{q}_{\mathrm{SP}})\mathbf{U}\mathbf{C}_{\mathrm{SP}}(\boldsymbol{q}_{\mathrm{SP}})^{\mathrm{T}}\Bigg/\sum_{\boldsymbol{q}_{\mathrm{SM}}}\mathbf{C}_{\mathrm{SM}}(\boldsymbol{q}_{\mathrm{SM}})\mathbf{U}\mathbf{C}_{\mathrm{SM}}(\boldsymbol{q}_{\mathrm{SM}})^{\mathrm{T}}}\tag{21}$$

4.2. Sound field feature matching for location and orientation detection

The location and orientation are determined by finding the maximum a posteriori location probability and a posteriori orientation probability for a given observation sequence:

$$\begin{aligned}\hat{l}&=\arg\max_{1\leqslant l\leqslant L}F_{\mathrm{RLM}}(l)\\&=\arg\max_{1\leqslant l\leqslant L}\alpha_{\mathrm{LP}}G(\boldsymbol{\lambda}_{\mathrm{LP}}(l)|\mathbf{P}_{\mathrm{LY}})+\alpha_{\mathrm{LM}}G(\boldsymbol{\lambda}_{\mathrm{LM}}(l)|\mathbf{M}_{\mathrm{LY}})\\&=\arg\max_{1\leqslant l\leqslant L}\alpha_{\mathrm{LP}}(G(\mathbf{P}_{\mathrm{LY}}|\boldsymbol{\lambda}_{\mathrm{LP}}(l))p(\boldsymbol{\lambda}_{\mathrm{LP}}(l))/p(\mathbf{P}_{\mathrm{LY}}))\\&\quad+\alpha_{\mathrm{LM}}(G(\mathbf{M}_{\mathrm{LY}}|\boldsymbol{\lambda}_{\mathrm{LM}}(l))p(\boldsymbol{\lambda}_{\mathrm{LM}}(l))/p(\mathbf{M}_{\mathrm{LY}}))\end{aligned}\tag{22}$$

$$\begin{aligned}\hat{o}&=\arg\max_{1\leqslant o\leqslant O}F_{\mathrm{ROM}}(o)\\&=\arg\max_{1\leqslant o\leqslant O}\alpha_{\mathrm{OP}}(G(\mathbf{P}_{\mathrm{OY}}|\boldsymbol{\lambda}_{\mathrm{OP}}(o))p(\boldsymbol{\lambda}_{\mathrm{OP}}(o))/p(\mathbf{P}_{\mathrm{OY}}))\\&\quad+\alpha_{\mathrm{OM}}(G(\mathbf{M}_{\mathrm{OY}}|\boldsymbol{\lambda}_{\mathrm{OM}}(o))p(\boldsymbol{\lambda}_{\mathrm{OM}}(o))/p(\mathbf{M}_{\mathrm{OY}}))\end{aligned}\tag{23}$$

where $\mathbf{P}_{\mathrm{SY}}=\left\{\boldsymbol{P}_{\mathrm{SY}}^{(1)},\ldots,\boldsymbol{P}_{\mathrm{SY}}^{(V)}\right\}$ and $\mathbf{M}_{\mathrm{SY}}=\left\{\boldsymbol{M}_{\mathrm{SY}}^{(1)},\ldots,\boldsymbol{M}_{\mathrm{SY}}^{(V)}\right\}$ are the phase difference and magnitude ratio computed from the testing sequences denoted as $Y_{\mathrm{S1}}(\omega)$ and $Y_{\mathrm{S2}}(\omega)$, and V denotes the testing sequence length. The probabilities $p(\boldsymbol{\lambda}_{\mathrm{LP}}(l))$ and $p(\boldsymbol{\lambda}_{\mathrm{LM}}(l))$ could be selected as $1/L$ and p $(\boldsymbol{\lambda}_{\mathrm{OP}}(o))$ and $p(\boldsymbol{\lambda}_{\mathrm{OM}}(o))$ could be selected as $1/O$ since the probability in each location and orientation is equally likely for a blind search. Moreover, because the probability densities $p(\mathbf{P}_{\mathrm{SY}})$ and $p(\mathbf{M}_{\mathrm{SY}})$ are the same for all location models, the detection rule can be recast as:

$$\hat{l}=\arg\max_{1\leqslant l\leqslant L}\alpha_{\mathrm{LP}}\prod_{v=1}^{V}G\left(\boldsymbol{P}_{\mathrm{LY}}^{(v)}|\boldsymbol{\lambda}_{\mathrm{LP}}(l)\right)+\alpha_{\mathrm{LM}}\prod_{v=1}^{V}G\left(\boldsymbol{M}_{\mathrm{LY}}^{(v)}|\boldsymbol{\lambda}_{\mathrm{LM}}(l)\right)\tag{24}$$

$$\hat{o}=\arg\max_{1\leqslant o\leqslant O}\alpha_{\mathrm{OP}}\prod_{v=1}^{V}G(\boldsymbol{P}_{\mathrm{OY}}^{(v)}|\boldsymbol{\lambda}_{\mathrm{OP}}(o))+\alpha_{\mathrm{OM}}\prod_{v=1}^{V}G\left(\boldsymbol{M}_{\mathrm{OY}}^{(v)}|\boldsymbol{\lambda}_{\mathrm{OM}}(o)\right)\tag{25}$$

5. Experimental results

Fig. 8 shows the experimental platform and the proposed RLODA. In Fig. 8a, the distance between two speakers is 0.2 m. Considering the spatial aliasing problem (Brandstein and Ward, 2001) and the highest frequency of sound generated by the robot, which is 2 kHz in this experiment, the distance between the two microphones of the RLODA is chosen as 0.07 m, as shown in Fig. 8b. The experiment was performed in an office room filled with furniture, which is 11.4 m in length, 4.73 m in width and 2.8 m in height. Two off-the-shelf, non-calibrated microphones are utilized on the ROLDA in this experiment and the RLODA is implemented on a PC with a stereo recording sound card. The sampling rate is 8 kHz, and the A/D resolution is 16 bit. The pre-recording is performed every 0.1 m within the region in which the robot is allowed to travel. For orientation detection, the robot is rotated in every 30° step to obtain 12 orientations in 360°.

Fig. 9 depicts the experimental environment and the location of the RLODA. Note that there is a partition room in the office. Therefore, the robot is completely under non-line-of-sight case when it is in the partition room. The robot's moving trajectories are also shown in Fig. 9 with the dotted lines from 1 to 8 in sequence.

The lengths of the training sequence and the testing sequence were set to 300 and 30. In other words, a three-second length input datum was set for training, and a 0.3 second length input datum was set for testing. The major noise in this experiment is speech noise and the minor noises are electric noise such as air conditioner noise, computer fan noise to simulate a general indoor environment.

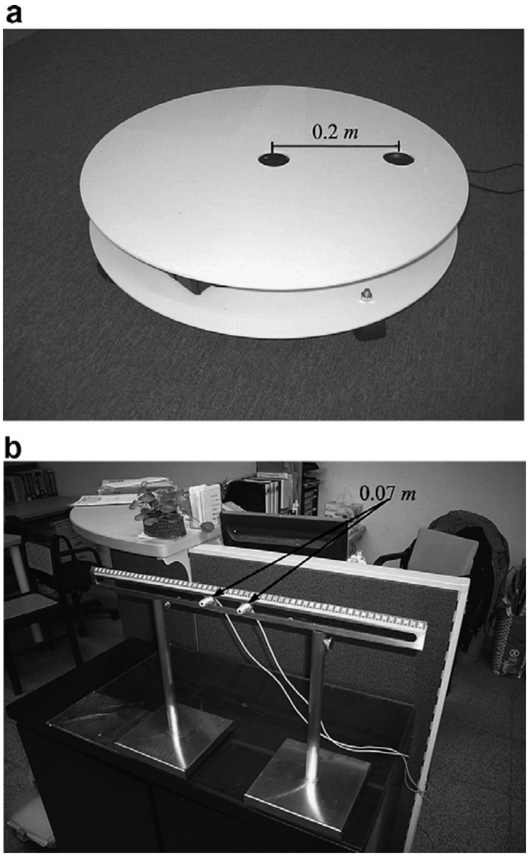

Fig. 8. The experimental platform and the proposed RLODA. (a) The experimental platform. (b) The proposed RLODA.

Table 1 lists the average SNRs of all trajectories and the average SNRs of each trajectory pair. Fig. 10 shows the location detection results along the robot's moving trajectory with a mixture number of 15 and an average SNR of 7.91 dB. As shown in Fig. 10, the location detection results are mostly very close to the actual location for most of the time.

The proposed method models the phase difference and magnitude ratio distributions measured from the sounds generated by the robot to perform robot's location and orientation detection. However, the sound field features of the noise start to dominate the phase difference and magnitude ratio distributions with the increment of noise power. In this circumstance, the RLMs and ROMs may become less distinguishable and may degrade the performance of the proposed method. In Fig. 10, the detection error occurs most frequently on trajectories 1 and 8, because some area of these trajectories is completely in the partition room and the average SNR of these trajectories is lower than those of other trajectories, as shown in Table 2. Although trajectories 1 and 8 contain locations that are in non-line-of-sight case, the location dependent sound field features can still be caught by the proposed RLMs.

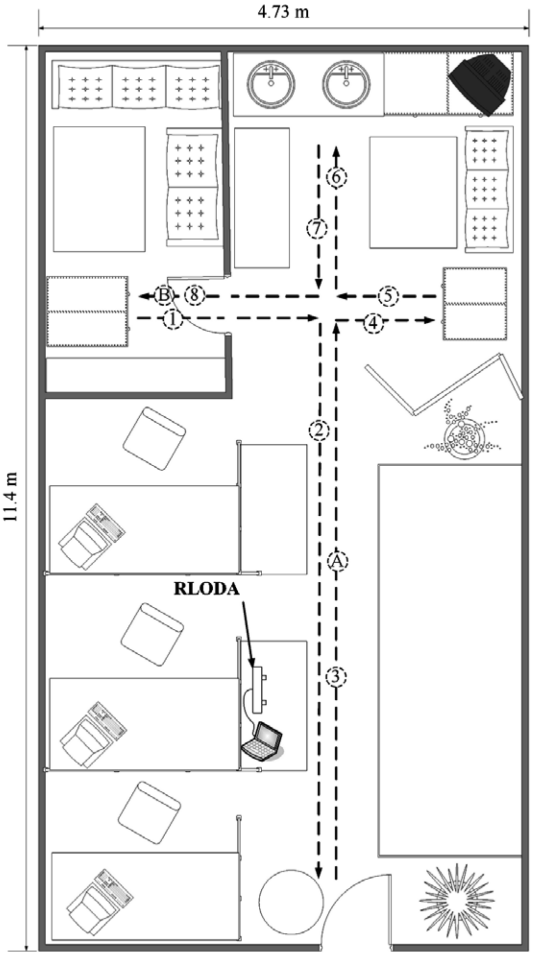

Fig. 9. Experimental environment.

Table 1
Average SNRs of all trajectories and the average SNRs of each trajectory pair (dB)

Average SNR	Average SNR of trajectories 1 and 8	Average SNR of trajectories 2 and 3	Average SNR of trajectories 4 and 5	Average SNR of trajectories 6 and 7
19.87	13.94	23.34	16.44	17.69
7.91	2.76	10.93	4.93	6.01

Several experiments are conducted to access the accuracy of the proposed method in terms of location and orientation detection error. Table 2 lists the average correct rates of the location detection results where D denotes the distance between the actual location and the nearest location in the pre-recorded database. Notably, the pre-recorded locations are discrete and are 0.1 m apart. In this experiment, if the detected result is the nearest pre-recorded location in the database, it will be regarded as a correct one. Additionally, the trial numbers for localization detection and orientation detection are 1210 and 332 individually for each condition. As shown in Table 2, if only a single Gaussian component is utilized ($M = 1$), then the average correct rates are too low to be acceptable in both two SNR cases. However, the average correct rates are improved to more than 95% when the mixture number is increased ($M = 11$ and $M = 15$) and $0 \leqslant D < 1$ cm.

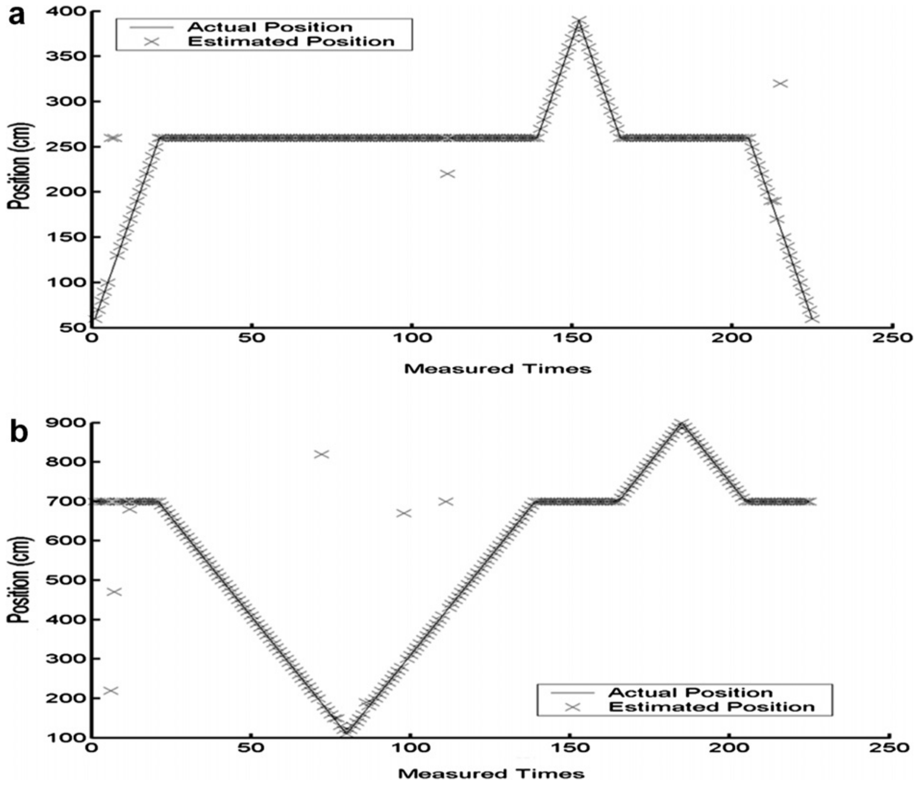

Fig. 10. Location detection results alone X and Y axes. (a) Location detection results alone X axis. (b) Location detection results alone Y axis.

Table 2
Average correct rates of location detection results (%)

Average SNR (dB)	$M = 1$			$M = 11$			$M = 15$		
	$0 \leqslant D < 1$ (cm)	$1 \leqslant D < 3$ (cm)	$3 \leqslant D < 5$ (cm)	$0 \leqslant D < 1$ (cm)	$1 \leqslant D < 3$ (cm)	$3 \leqslant D < 5$ (cm)	$0 \leqslant D < 1$ (cm)	$1 \leqslant D < 3$ (cm)	$3 \leqslant D < 5$ (cm)
19.87	24.00	20.83	20.41	95.45	95.00	85.45	97.19	95.00	88.35
7.91	22.98	22.89	17.52	91.98	89.50	84.13	94.38	87.93	81.57

Table 3 shows the average correct rates of the orientation detection results, where A denotes the distance between the actual and the pre-recorded orientations. If the orientation detection result is the nearest pre-recorded orientation to the actual orientation, the result will be considered correct. Note that the experiment is performed after a correct location is detected. As shown in Table 3, when $M = 1$, the average correct rates are lower than 60%. These results show that a single Gaussian component is not appropriate for modeling the ROMs. When $M = 11$, the average correct rates are much higher than those when $M = 1$ in both the SNR cases. In the condition of $0° \leqslant A < 4°$, the average correct rates exceed 99% in both the SNR cases.

Fig. 11 shows the average of a posteriori probabilities measured at the locations "A" and "B", where the location "A" is in a line-of-sight case and the location "B" is in a non-line-of-sight case, as illustrated in Fig. 9. Notably, the a posteriori location probability is defined as:

$$\alpha_{\mathrm{LP}} \prod_{v=1}^{V} G\left(\boldsymbol{P}_{\mathrm{LY}}^{(v)} | \boldsymbol{\lambda}_{\mathrm{LP}}(l)\right) + \alpha_{\mathrm{LM}} \prod_{v=1}^{V} G\left(\boldsymbol{M}_{\mathrm{LY}}^{(v)} | \boldsymbol{\lambda}_{\mathrm{LM}}(l)\right) \tag{26}$$

and the a posteriori orientation probability is defined as:

$$\alpha_{\mathrm{OP}} \prod_{v=1}^{V} G\left(\boldsymbol{P}_{\mathrm{OY}}^{(v)} | \boldsymbol{\lambda}_{\mathrm{OP}}(o)\right) + \alpha_{\mathrm{OM}} \prod_{v=1}^{V} G\left(\boldsymbol{M}_{\mathrm{OY}}^{(v)} | \boldsymbol{\lambda}_{\mathrm{OM}}(o)\right) \tag{27}$$

Table 3
Average correct rates of orientation detection results (%)

Average SNR (dB)	$M = 1$				$M = 11$			
	$0° \leqslant A < 4°$	$4° \leqslant A < 8°$	$8° \leqslant A < 12°$	$12° \leqslant A < 15°$	$0° \leqslant A < 4°$	$4° \leqslant A < 8°$	$8° \leqslant A < 12°$	$12° \leqslant A < 15°$
19.16	58.43	48.49	45.78	44.28	99.70	88.55	84.04	81.33
7.39	58.13	50.00	50.00	48.19	99.10	84.34	80.12	77.11

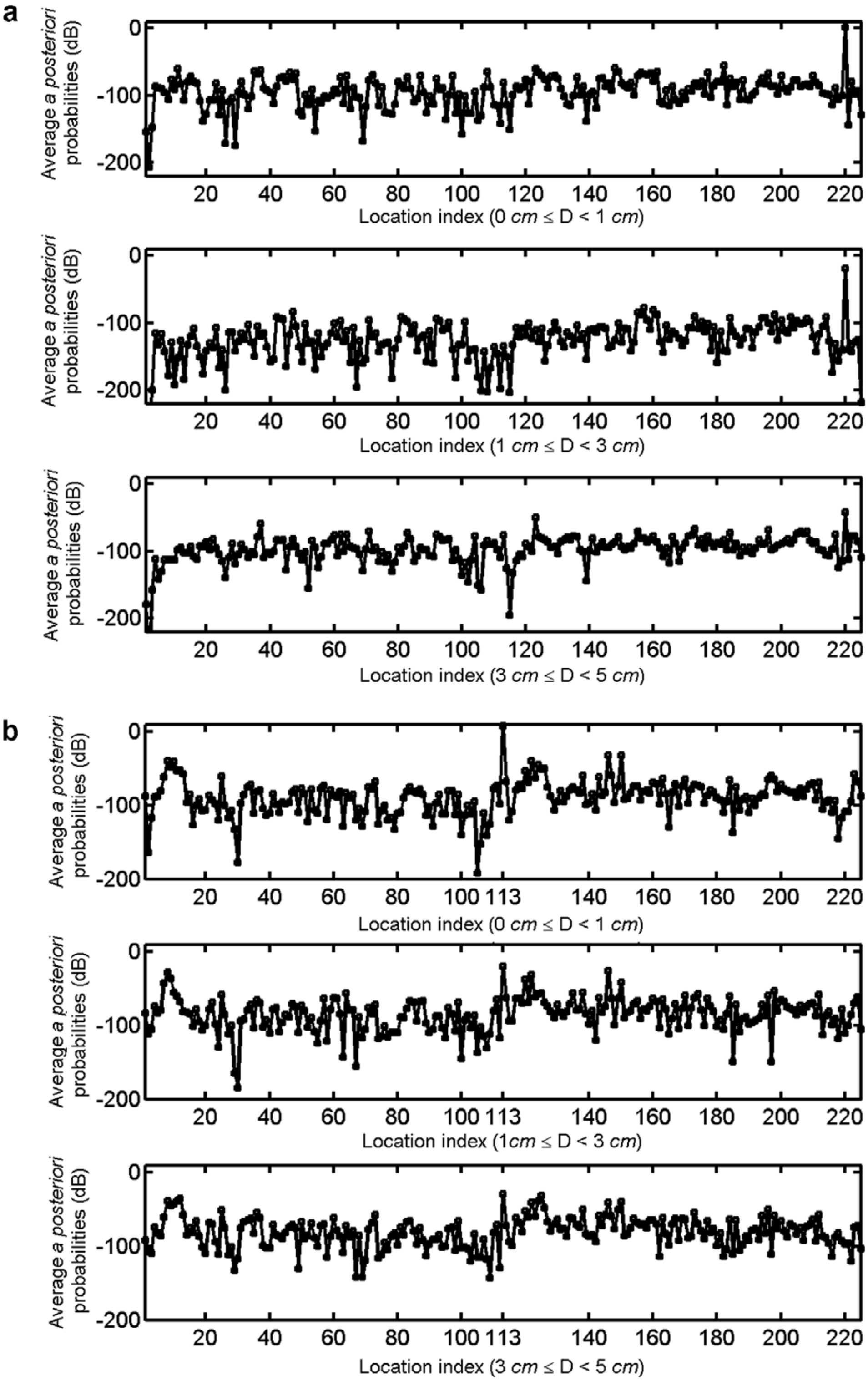

Fig. 11. The average of the measured a posteriori probabilities. (a) The average a posteriori location probabilities at the location “A”. (b) The average a posteriori location probabilities at the location “B”. (c) The average a posteriori orientation probabilities at the location “A”. (d) The average a posteriori orientation probabilities at the location “B”.

The average SNRs belong to the lowest SNR conditions in Tables 2 and 3 individually. The mixture number in Fig. 11a and b is 15, and the mixture number in Fig. 11c and d is 11. The location “A” denotes the 113th location and the location “B” represents the 220th location. In the case of $0 \leqslant D < 1$ cm, the averages of (26) (averages of a posteriori location probabilities) measured with the correct locations indices ($l = 113$ and 220) are much higher than those of other location indices, as shown in Fig. 11a and b. However, since the sound field feature varies with the ro-

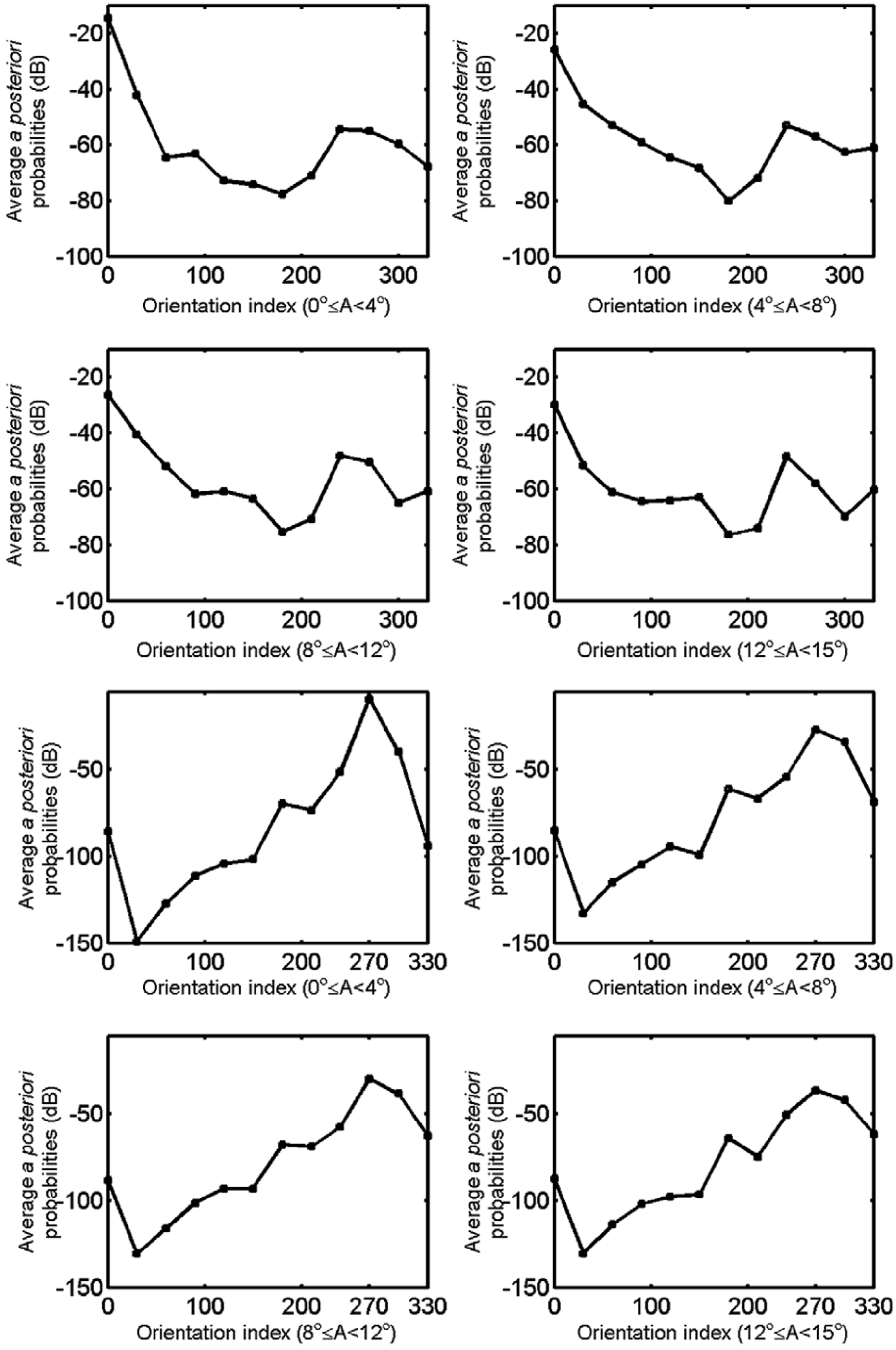

Fig. 11 (*continued*)

bot's location and orientation, the phase difference and magnitude ratio distributions are becoming less similar while the robot is moving away from the pre-recorded location or orientation. Therefore, in Fig. 11a and b, the difference between the averages of (26) measured with the correct locations indices and with other location indices are becoming less obvious with the increase of D, and then the chance of detection error rises. This tendency explains why the average correct rates of location detection in Table 2 degrade with the increase of the distances between the actual and the pre-recorded locations. Although the averages of (26) measured with the correct locations indices decrease with the increase of D, it is still higher than those measured with other location indices; as a result, the correct rates listed in Table 2 remain above 80% when $3 \leqslant D < 5$ cm. The same phenomenon appears in the experiment of orientation detection. Fig. 11c and d depicts the average of (27) (averages of a posteriori orientation probabilities) with the correct orientations of 0° for Fig. 11c and 270° for Fig. 11d. The average of (27) measured at the correct orientation indices drops with the increase of A in both line-of-sight and non-line-of-sight cases and so does the average correct rates of the orientation detection in Table 3. These experimental results show that utilizing GMMs to model the sound field features is a feasible method for robot's location and orientation detection.

6. Conclusion

A novel robot's location and orientation detection method based on sound field features matching is proposed. The proposed method treats phase difference and magnitude ratio distributions between the microphones as distinct sound field features, and models them by GMMs to detect a robot's location and orientation. Since the proposed method makes no assumptions about the spatial relationship between sound sources and microphones, it can be applied to both line-of-sight and non-line-of-sight cases. Moreover, the modeled sound field features are content independent, so the content of sound can be designed arbitrarily. A system architecture is also proposed to provide robustness to environmental noises. The proposed method is suitable to be integrated with other robot location or orientation detection algorithms based on different sensors to provide initial conditions for reducing the search effort, or to compensate for localizing certain locations that cannot be detected using other localization methods to perform more robust, more accurate and faster pose and global location detection.

Acknowledgements

This work was supported by National Science Council of the ROC under Grant No. NSC94-2218-E009064 and MOE ATU Program under the account number 95W803E.

References

Argamon-Engelson, S., 1998. Using image signatures for place recognition. Pattern Recog. Lett. 19 (10), 941–951.

Borenstein, J., Everett, H.R., Feng, L., 1996. Navigating Mobile Robots: Sensors and Techniques. A.K. Peters, Wellesley, MA.

Brandstein, M.S., Silverman, H.F., 1997. A robust method for speech signal time-delay estimation in reverberant rooms. IEEE Int. Conf. Acoust. Speech Signal Process. 1, 375–378.

Brandstein, M., Ward, D., 2001. Microphone Arrays: Signal Processing Techniques and Applications. Springer-Verlag, New York, p. 26 (Chapter 2).

Carter, G.C., Nuttall, A.H., Cable, P.G., 1973. The smoothed coherence transform. IEEE Sig. Process. Lett. 61, 1497–1498.

Elkan, C., 2003. Using the triangle inequality to accelerate *k*-means. Proc. 20th Int. Conf. Machine Learning, 147–153.

Georgiev, A., Allen, P.K., 2004. Localization methods for a mobile robot in urban environments. IEEE Trans. Robot. 21, 851–864.

Gutierrez-Osuna, R., Janet, J.A., Luo, R.C., 1998. Modeling of ultrasonic range sensors for localization of autonomous mobile robots. IEEE Trans. Ind. Electron. 45 (4), 654–662.

Knapp, C.H., Carter, G.C., 1976. The generalized correlation method for estimation of time delay. IEEE Trans. Acoust. Speech Signal Process. 24, 320–327.

Ladd, A.M., Bekris, K.E., Rudys, A.P., Wallach, D.S., Kavraki, L.E., 2004. On the feasibility of using wireless ethernet for indoor localization. IEEE Trans. Robot. Automat. 20 (3), 555–559.

Larsson, U., Frosberg, J., Wernersson, A., 1996. Mobile robot localization: Integrating measurements from a time-of-flight laser. IEEE Trans. Ind. Electron. 43 (3), 422–431.

Lee, J.M., Son, K., Lee, M.C., Choi, J.W., Han, S.H., Lee, M.H., 2003. Localization of a mobile robot using the image of a moving robot. IEEE Trans. Ind. Electron. 50 (3), 612–619.

MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. Proc. Fifth Berkeley Symp. Mathematical Statistics and Probability, 281–297.

McGillem, C.D., Rappaport, T.S., 1988. Infra-red location system for navigation of autonomous vehicles. IEEE Int. Conf. Robot. Automat., 1236–1238.

Nikas, C.L., Shao, M., 1995. Signal Processing with Alpha-Stable Distributions and Applications. Wiley, New York.

Ohya, I., Kosaka, A., Kak, A., 1998. Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing. IEEE Trans. Robot. Automat. 14 (6), 969–978.

Parker, S.P., 1988. Acoustic Source Book. McGraw-Hill, New York.

Reynolds, D.A., Rose, R.C., 1995. Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. Speech Audio Process. 3 (1), 72–83.

Tamai, Y., Kagami, S., Mizoguchi, H., Amemiya, Y., Nagashima, K., Takano, T., 2004a. Real-time two-dimensional sound source localization by 128-channel huge microphone array. IEEE Int. Workshop Robot Human Interact. Commun., 65–70.

Tamai, Y., Kagami, S., Mizoguchi, H., Amemiya, Y., Nagashima, K., Takano, T., 2004b. Sound spot generation by 128-channel surround speaker array. IEEE Int. Workshop Sensor Array Multichannel Sig. Process., 542–546.

Vlassis, N., Motomurat, Y., Hara, I., Asoh, H., 2001. Edge-based features from omnidirectional images for robot. Proc. IEEE Int. Conf. Robot. Automat., 1579–1584.

Wang, Q.H., Ivanov, T., Aarabi, P., 2004. Acoustic robot navigation using distributed microphone arrays. Inform. Fusion 5, 131–140.

Weiss, G., Wetzler, C., von Puttkamer, E., 1994. Keeping track of position and orientation of moving indoor systems by correlation of range-finder scans. Proc. IEEE/RSJ/GI Int. Conf. IROS, 595–601.

Xuan, G., Zhang, W., Chai, P., 2001. EM algorithms of Gaussian mixture model and hidden Markov model. IEEE Int. Conf. Image Process., 145–148.

Yamada, M., Itsuki, N., Kinouchi, Y., 2004. Adaptive directivity control of speaker array. Control Automat. Robot. Vis. Conf., 1143–1148.