國立交通大學

電機與控制工程學系

博士論文

雙耳特徵差異分佈模版於非靜態聲源之定位研究

Binaural room distribution pattern for nonstationary sound source localization

研 究 生：劉維瀚

指導教授：胡竹生 教授

中華民國九十六年九月

雙耳特徵差異分佈模版於非靜態聲源之定位研究

Binaural room distribution pattern for nonstationary sound source localization

研究生：劉維瀚　　Student：Wei-Han Liu

指導教授：胡竹生　　Advisor：Jwu-Sheng Hu

國立交通大學

電機與控制工程學系

博士論文

A Dissertation

Submitted to Department of Electrical and Control Engineering

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Electrical and Control Engineering
September 2007
Hsinchu, Taiwan, Republic of China

中華民國九十六年九月

雙耳特徵差異分佈模版於非靜態聲源之定位研究

研究生：劉維瀚　　　　　　　　　　指導教授：胡竹生 博士

國立交通大學電機與控制工程學系(研究所)博士班

摘要

在現實聲源定位的應用環境中，自然聲源之統計特性通常為非靜態（nonstationary），而環境則會造成複雜的迴響（reverberation）。因此，非靜態聲源於迴響環境中之定位，即成為工程學上重要的研究議題。本篇論文探討非靜態聲源與雙耳特徵差異（IPD、ILD）間之關係。在本篇論文中，採用移動極點模型的概念，提出以指數多項式建立非靜態聲源的強度波動模型。根據此模型，本論文提出利用 IPD、ILD 的分佈模版做為聲源定位之充分條件，並解釋分佈模版中多重峰值出現出原因。此外，本論文亦提出以高斯混合模型為基礎之「高斯雙耳特徵差異分佈模型」(GMBRDM)，作為非靜態聲源定位之演算法。此部分所提出之理論與演算法，皆有模擬或實驗結果加以討論與驗證。

除此之外，本論文將研究之非靜態聲源定位之方法應用於機器人室內定位環境，提出一創新之機器人位置與方向偵測系統。此系統適用於迴響複雜度高之環境，並具有對雜訊穩健之特性。實驗結果顯示，本系統可以用於近場與遠場環境，亦可在機器人與麥克風間無直接傳導路徑時使用。由於本系統可以執行機器人之全域定位，因此適合與其他定位方式整合，作為提供初始化參數或補償之用。

Binaural room distribution pattern for nonstationary sound source localization

Graduate Student: Wei-Han Liu Advisor: Dr. Jwu-Sheng Hu

Department of Electrical and Control Engineering
National Chiao-Tung University

Abstract

Nature sound sources are usually nonstationary and the real environment contains complex reverberations. Therefore, nonstationary sound source localization in a reverberant environment is an important research topic. This dissertation discusses the relationships between the nonstationarity of sound sources and the distribution patterns of interaural phase differences (IPDs) and interaural level differences (ILDs) based on short-term frequency analysis. The level fluctuation of nonstationary sound sources is modeled by the exponent of polynomials from the concept of moving pole model. According to this model, the sufficient condition for utilizing the distribution patterns of IPDs and ILDs to localize a nonstationary sound source is suggested and the phenomena of multiple peaks in the distribution pattern can be explained. Simulation is performed to verify the proposed analysis. Furthermore, a Gaussian-mixture binaural room distribution model (GMBRDM) is proposed to model distribution patterns of IPDs and ILDs for nonstationary sound source localization. The effectiveness and performance of the proposed GMBRDM are demonstrated by experimental results.

The proposed nonstationary sound source localization algorithm is adopted for robot localization application. A novel and robust robot location and orientation detection method based on sound field features is proposed. Unlike conventional methods, the proposed method does not explicitly utilize the information of direct sound propagation path from sound source to microphones, nor attempt to suppress the reverberation and noise signals. Instead, the proposed method utilizes the sound field features obtained when the robot is at different location and orientation in an indoor environment. The experimental results show that the proposed method using only two microphones can detect robot's location and orientation under both line-of-sight and non-line-of-sight cases and can be applied to both near-field and far-field conditions. Since this method can provide global location and orientation detection, it is suitable to fuse with other localization methods to provide initial conditions for reduction of the search effort, or to provide the compensation for localizing certain locations that cannot be detected using other localization methods.

致 謝

首先要感謝我的指導教授胡竹生教授之指導，在博士班的學習中，最重要的就是指導教授對學問的態度以及寶貴的知識。胡教授不僅對學問富有興趣，也對研究抱有極大的熱誠。在研究過程中，更是給了我很大的思考與發揮的空間。除此之外，老師也安排我參加各種競賽與計畫，更是讓我得到許多實作的經驗。

感謝我的父母親的栽培，我知道你們很辛苦，沒有你們的支持，我也沒有辦法順利畢業。同時也要感謝我親愛的老婆，還好妳在這段時間裡面不離不棄，我才能無後顧之憂的進行我的學業。

感謝實驗室眾多的學長、同學、以及學弟，你們不但給了我很多研究上的幫助，也讓漫長的研究過程中增添了繽紛的色彩。余祥華學長、蕭得聖學長，謝謝你們在研究上給我的建議，使我的學習過程更為順利。rrrr 學長、凱學長，感謝你們把 TI DSP 的技術灌頂給我，成為我日後混飯吃的一大利器。小陶子學長對鋼彈、EVA 和其他老動畫的瞭解至今無人能及，對作業系統與程式設計的深厚知識更是實驗室所有同學耳中的傳奇。還有阿邦、瓊宏、俊德學長，你們帶領我進入 XLAB，並建立起良好的實驗室環境。立偉學長，你的投影片長度目前仍然坐穩實驗室第一的紀錄。伯彥學長，你的 FTP 一度成為我的精神支柱，陪我度過不少苦悶的日子。

感謝鄭博士，在這段時間裡與我一起做研究、接計畫、寫論文，我們在車上錄音的經驗真是令人畢生難忘。還有從高中就一直同學到現在的宗敏，我最懷念我們一起討論 Star Trek 的日子了。實驗室第一位女性學生－欣慈，妳讓我充分體認什麼叫做巾幗不讓鬚眉，「我在懈怠」一語更是如雷貫耳，成為大家佩服的對象。感謝俊葦讓我搞懂小畫家到底對寫報告有什麼幫助，你充滿創意與實用性的用法對我助益良多。還有育德，我們一同設計數位麥克風電路與做實驗的日子

也是我博士班生涯中重要的一環，雖然後來我沒有辦法與你一同創造多頭，但還是感謝你帶我參觀了 CIC 優良的工作環境與昂貴的機台。

感謝 Alan、昊群、Pazz、Angel、還有鏗兄、弘齡、楷祥，你們在這段時間裡面一直陪我去重訓，真是我難忘的回憶。鏗兄，以後重訓組就交給你了！還有實驗室最 man 的興哥，還好有你的麥克風陣列平台，我們才能輕鬆的完成各種計畫。感謝嘉芳、德琪、憶如、岑思與佩靜、鳥蕙、瓊文，你們不但讓實驗室增添許多不同的氣氛，而且對動畫也有頗深的造詣。

感謝攝影組的鳥哥、Alpha、以及恆嘉，你們對我的建議讓我對攝影這門藝術有了初步的瞭解，還好我有忍住，不然就落入 DSLR 的無底深淵了。另外還要感謝與我一同做實驗的 Papa、永融，沒有你們的幫忙與合作，實驗的困難度會大幅提昇。還有常鬥嘴的倉億、阿鍇、吃素的俊德、跟原住民混最熟的家瑋、實作能力超強的康康、有布掛在身上就不會冷的群祺、為了 ITS 考駕照的晏容、腸胃不好的士奇、把妹一流的豬木、流行感十足的耀賢、總是話不多的螞蟻、最有博士相的阿吉、可愛的俊宇、照相 pose 一流的 Gun、外號比本名難念的 HCY，新進實驗室的育綸、畢業生代表嘟嘟、開車超猛的唐哥，你們的陪伴讓實驗室充滿歡樂。

最後，也是最重要的，感謝我的主耶穌，感謝祢垂聽我的禱告，引領我度過各種困難，保守我走在平安的路上。

Contents

Index

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Sound Source Localization Using Binaural Information

The task of localizing a sound source using multiple microphones has been developed for years [1]. Among various kinds of techniques, methods that are based on the auditory system of humans or other animals using two microphones are one of the most popular approaches in this research field.

Sound perceived by humans is influenced by the torso, pinna, shape of human head, and acoustic environment. To identify how the human body affects perceived sound waveforms, head-related transfer function (HRTF), or head-related impulse response (HRIR) were proposed [2 - 4]. Generally, HRIR is a measure of impulse response from the sound source to eardrums in an anechoic room [5]—HRTF is considered the Fourier transform of HRIR [6]. Since HRTF varies with the sound source location, many localization cues were discussed based on HRTF. For example, the interaural level differences (ILDs) and the interaural time differences (ITDs) are utilized as the major cues for localizing a sound source, especially for azimuth localization [7].

1.1.1 Azimuth Localization Using Binaural Localization Cues

Brungart *et al*. concluded that ILDs play an important role in localizing sound sources near the head [8]. The ITD, or interaural phase difference (IPD), which is the frequency domain representation of ITD, can be estimated by cross-correlation functions [9] or generalized cross-correlation (GCC) methods [10, 11]. Although IPDs and ILDs have been studied for a long time, they have limitations. Based on an assumption that head and ears are symmetrical, the sound source presented at a median plane should produce no interaural difference. Therefore, interaural difference cues are insufficient for localizing the elevation of a sound source in the medium plane. Moreover, any sound source falls on a “cone of confusion,” as Woodworth called it [12], may lead to constant IPDs or ILDs.

1.1.2 Elevation Localization Using Binaural Localization Cues

Cues of spectral modification are very important for elevation localization and front-back discrimination [13]. As the sound is filtered by the pinna before reaching the eardrum, “pinna notch” was investigated [14]. The influence of head diffraction and torso reflection was also examined [15]. Therefore, the elevation of a sound source can be estimated by comparing the incoming spectrum with the stored HRTF [16, 17]. These elevation estimation methods typically assume a flat sound spectrum or one that is known in advance. In practice, natural sounds are highly nonstationary and localization systems have no *a priori* knowledge of the spectrum shape of the nature sound [16]. Although most research results showed that spectral modification cues are significant for elevation localization, it remains unclear how the human auditory system localizes the elevation of nonstationary sound [16].

1.1.3 Distance Localization Using Binaural Localization Cues

Another significant localization ability of the human auditory system is distance localization. Research works indicated that many possible cues exist for distance localization (e.g., overall sound source intensity and energy ratio of direct to reverberant sound [7] [18]). However, overall sound source intensity can only be employed for relative distance localization and the energy ratio of direct to reverberant sound is strongly influenced by the reflections within an application environment [7]. Therefore, sound source localization in three-dimensional environments using binaural information remains an open research topic.

1.2 An Overview of Microphone-Array-Based Direction of Arrival Estimation

Besides HRTF based approaches, methods based on multiple microphones or microphone array are proposed. Figure 1-1 shows the physical layout of a uniform linear microphone array.

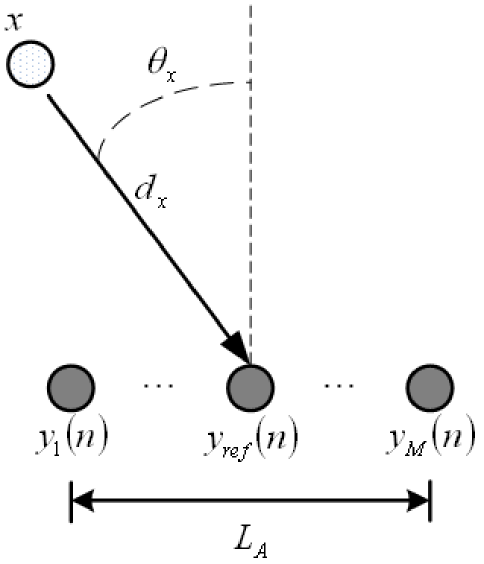


Figure 1-1 Physical layout of a uniform linear microphone array.

where x denotes the sound source, $y_1(n)\cdots y_M(n)$ denotes the signal received by M microphones, $y_{ref}(n)$ denotes the signal received by the reference microphone, L_A is the size of the microphone array and d_x, θ_x are the distance and azimuth from the sound source to the reference microphone. Based on the relation between L_A and d_x, the received signals can be regard as plane waves (far-field) or spherical waves (near-field). The definition of far-field and near field can be found in [19].

Generally, existing microphone-array-based sound source localization algorithms may be divided into three categories: steer-beamformer-based algorithms, eigen-structure-based direction of arrival (DOA) estimation algorithms, and time-delay of arrival (TDOA) based algorithms.

1.2.1 Steer-Beamformer-Based Algorithms

Steer-beamformer-based sound source localization algorithms [20 - 22] utilize beamformer algorithms to form beams for spatial filtering and steer the formed beam to the interested directions to obtain the spatial response. The power of the spatial response is then computed and the most possible sound source direction is decided by finding the direction with maximum power. The performance of steer-beamformer-based algorithm depends mainly on the resolution of the beamformer and the steer-beam algorithm adopted. Therefore, the performance of steer-beamformer-based sound source localization algorithms is limited by the resolution of beamformer. Raising the resolution of beamformer and increasing the steered beam direction result in higher computational load.

1.2.2 Eigen-Structure-Based DOA Estimation Algorithms

The eigen-structure-based DOA estimation algorithms [23 - 34] are proposed for high-resolution multiple sound source localization. This kind of sound source localization algorithms derive the data correlation matrix from the signals obtained from the microphone array and compute the eigenvectors. The eigenvectors are separated into two subspaces, signal subspace and noise subspace, according to the importance of eigenvalue. Based on the theory in [24], the array manifold vector (or the steering vector), $a(\theta)$, corresponds to the sound source localization and is orthogonal to the noise subspace. Consequently, the projection of $a(\theta)$ to the noise subspace must be zero theoretically when θ consists the direction of sound source.

Figure 1-2 is a three-dimensional example of the relation between signal subspace, noise subspace and manifold vectors, where E_1 and E_2 are the eigenvectors that span signal subspace, and E_3 is the eigenvector that span noise subspace. Therefore, the manifold vector $a(\theta)$ is orthogonal to E_3 when $\theta = \theta_1$ or θ_2. According to this relation, the location of sound sources can be estimated. The drawbacks of eigen-structure-based DOA estimation algorithms are the ability of dealing with reverberant signal and the requirement of eigenvalue decomposition. When the environment is reverberant, the eigen-structure of the data correlation matrix would break the assumption above and result in an un-robust estimation result. On the other hand, the requirement of eigenvalue decomposition makes eigen-structure-based DOA estimation algorithms need more processing power.

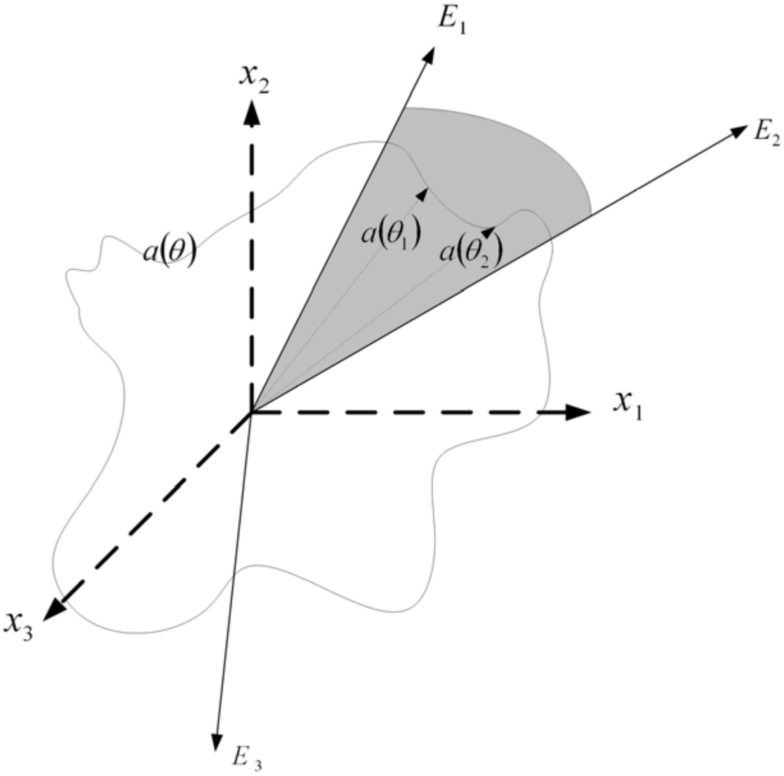


Figure 1-2 An illustration of signal subspace, noise subspace and manifold vectors.

1.2.3 Time-Delay of Arrival Based Algorithms

The TDOA based algorithms [35 - 39, 10, 11] measure the time delay using phase difference between microphone pairs. Cross-correlation methods and GCC methods discussed in 1.1.1 can also be classified in this category. The time delay is then combined with the geometric relation between sound source and microphones to estimate the most possible sound source location. Basically, TODA based algorithms require lower computation power then methods in the other two categories. However, TODA based algorithms are sensitive to the weights of processed frequencies. Works on how to select a set of optimal weight are proposed and discussed in this research field.

1.3 Known Problems in Sound Source Localization

Early experimental results for HRTF were principally obtained in anechoic rooms using maximum-length sequence (MLS) method [40]. Since this approach is based on stationary sound sources in anechoic rooms, conventional studies of HRTF mainly concentrated on the steady-state response from a sound source to the eardrums caused by human body. Only a few studies addressed the issue of localizing a sound source in a reverberant room [14] [41]. In a real enclosure, the relation between a sound source and microphones is very complicated and is almost impossible to characterize with a finite-length data and short-term analysis method such as short-term Fourier transform (STFT). According to the investigation of room acoustics [42], the number of eigen-frequencies with an upper limit of $f_s/2\,\text{Hz}$ can be obtained by the following equation:

$$L = \frac{4\pi}{3}\Psi\left(\frac{f_s}{2\nu}\right)^3 \qquad (1\text{-}1)$$

where f_s denotes the sampling frequency, ν represents the sound velocity ($\nu \approx 340\,\text{m/s}$) and Ψ is the geometrical volume. This equation indicates that the number of poles is too high when the frequency is high, and that the transient response occurs in almost any processing duration when the input signal is a nonstationary sound. For example, the number of poles is about 96435 when the sampling frequency is 8000 Hz and the volume is 14.1385 m^3. Hence, the nonstationary characteristic of nature sound source makes the IPDs and ILDs between the signals received by two microphones from a fixed sound source vary among data sets.

In practice, reverberant sounds can significantly influence the localization cues. Gustafsson *et al*. analyzed how reverberation can distort time-delay estimation [43]. Shinn-Cunningham *et al*. showed that HRTFs are altered by reverberant sound in a classroom [44, 45] and the reverberation can cause temporal fluctuation in short-term IPDs and ILDs [44]. These studies suggest that the performance of general methods of sound source localization based on a set of HRTFs measured in anechoic rooms with stationary sound sources can be limited because of the nonstationary property of natural sound, reverberation, and short-term frequency analysis. Based on the precedence effect, some sound source localization methods excluded reverberant sound by detecting sound source onset [46]. However, as mentioned above, reverberation actually helps listeners judge the distance of sound source. Excluding reverberant sound can restrict the ability for distance localization.

Besides reverberation and nonstationarity of sound source, there are still other non-ideal issues such as the non-line-of sight condition and microphone mismatch problem. When only two microphones are used, the methods mentioned above estimate mainly the difference between microphones. It means that the methods cannot distinguish between different sound sources, which are aligned relative to the array under far-field condition. Furthermore, barriers may exist between microphones and sound source (so-called the non-line-of sight condition) 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, 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 characteristic of microphone changes with sound direction and is hard to measure precisely.

1.4 Contribution of this Dissertation

This work focuses on localizing sound source in complex indoor environment. Unlike traditional works that try to eliminate the influence of the temporal fluctuation caused by reverberations, this work attempts to model these fluctuations using statistical models for sound source localization.

In the first part of this work, the relation between the nonstationarity of sound source and the distribution patterns of IPDs and ILDs when short-term frequency analysis is utilized for analysis is discussed. The level fluctuation of nonstationary sound source is modeled by the exponential of polynomials based on the concept of moving pole model. Accordingly, the sufficient condition for utilizing the distribution patterns of IPDs and ILDs to detect the location of nonstationary sound source is suggested and the phenomena of multiple peaks in the distribution patterns can also be explained. Furthermore, a Gaussian mixture based model, called Gaussian-mixture binaural room distribution model (GMBRDM) is proposed to model the distribution patterns of IPDs and ILDs for nonstationary sound source localization.

In the second part, the related research is applied to robot’s location and orientation detection. 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 (GMMs) [53] are utilized in this work 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 DOA estimation using propagation delay. They usually suffer from reverberation, non-line-of-sight and microphone mismatch problems.

1.5 Organization of this Dissertation

This dissertation is organized as follows. This chapter provides a brief introduction to the general sound source localization algorithms, including methods which follow the auditory system of human and methods based on microphone array. Moreover, this chapter also discusses the main contribution and the organization of this dissertation. In the next chapter, the binaural room distribution pattern and related GMBRDM are introduced. In Chapter 3, sound based robot's location and orientation detection system is proposed and discussed. The related experiments are shown in Chapter 4. Chapter 5 gives some conclusion remarks and avenues for future research.

Chapter 2

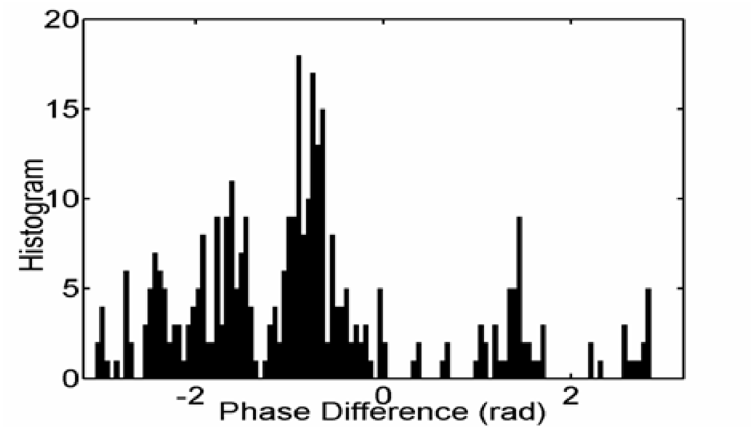
Nonstationary Sound Source Localization Using Binaural Room Distribution Pattern

2.1 Introduction

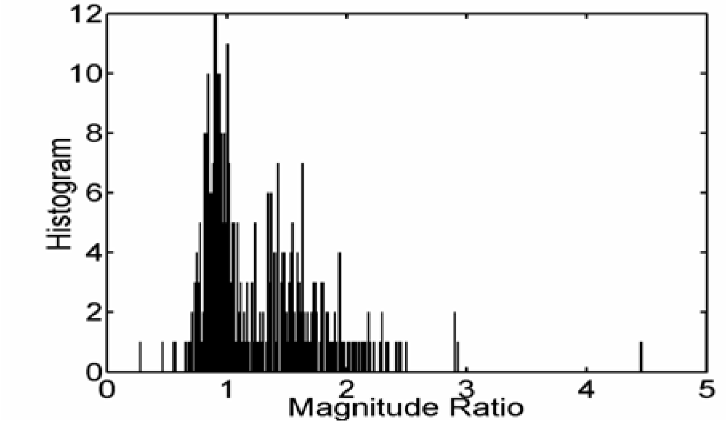
As discussed in the first chapter, localizing a nonstationary sound source in a reverberant environment can face temporal fluctuation of interaural cues. Recent research results for sound source localization revealed the importance of temporal fluctuation phenomenon of IPDs and ILDs. Rather than eliminate the influence of these fluctuations, these studies attempted to describe these fluctuations using statistical models for sound source localization [56]. The work in [48] investigated localization cues of IPDs and ILDs exhibiting temporal fluctuation phenomena when sound sources are nonstationary and short-term frequency analysis, such as short-term Fourier transform (STFT), is utilized. In [48], distribution patterns of IPDs and ILDs were calculated from the superposition of sound sources recorded in an anechoic room and spatially distributed noise recorded in real environments. The distribution patterns were applied to estimate the azimuth and elevation of a sound source using Bayesian maximum *a posteriori* estimation. The experiments demonstrated good results in both quiet and noisy conditions. Further, Smaragdis and Boufounos [49] also used the Gaussian model to model the empirical features in a reverberant room. In their work, the fluctuation of relative magnitude and phase of the cross spectra is modeled for sound source localization and the wrapping effect is solved using the proposed wrapped Gaussian model.

This study attempts to probe further the cause of IPD and ILD distribution patterns when a sound source is nonstationary and STFT is utilized. To simplify the description, distribution patterns of IPDs and ILDs are called binaural room distribution patterns (BRDPs) in the remainder of this work. The idea of moving pole model is employed to model the nonstationary sound sources; consequently, the level fluctuation is modeled as an exponent of polynomial. Based on this model, it can be shown that BRDPs depend on the content of the nonstationary source signals. The dependency is analyzed to explain the phenomenon of multiple peaks in the BRDPs.

In real environment, more than one peak can exist in the measured BRDPs. For example, Fig. 2-1 illustrates the IPDs and ILDs measured at the location marked “A” in Fig. 2-2.



(a)



(b)

Figure 2-1 The histograms of IPDs and ILDs measured at location the marked “A”. (a) The histogram of IPDs at the location marked “A”. (b) The histogram of ILDs at the location marked “A”.

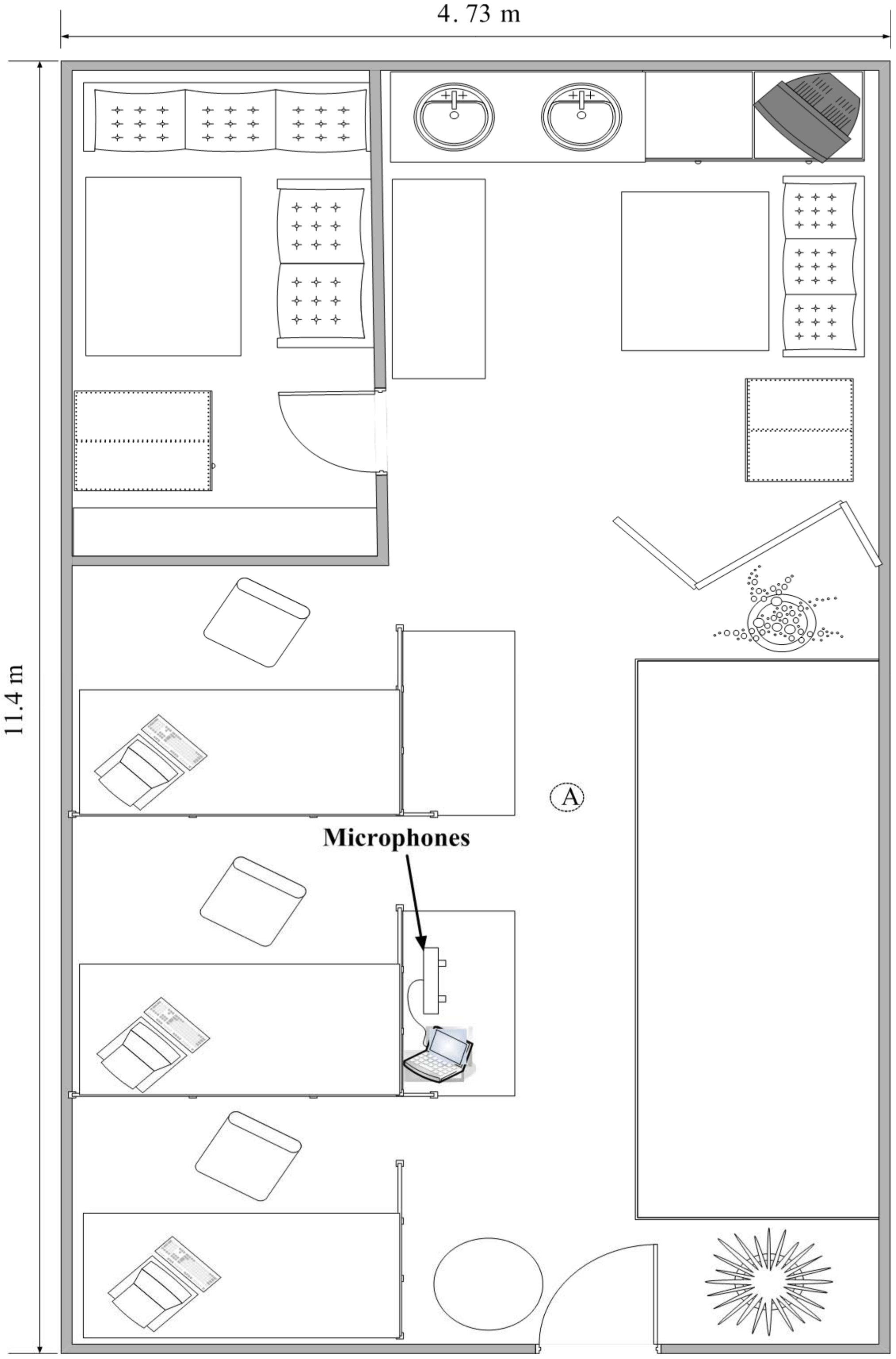


Figure 2-2 The recording environment.

As shown in Fig. 2-1, the IPD and ILD contain multiple peaks. This phenomenon can be explained with the proposed model.

Since the BRDPs can contain multiple peaks, a modeling method that deals with complicated distribution patterns is needed. Although the work in [48] utilized normalized histograms to model distribution patterns, the memory requirement is considerable when histogram resolution is high. Therefore, this work adopts GMMs to model BRDPs and proposes a GMBRDM to parameterize them. Because the proposed GMBRDM is a linear combination of the phase difference GMM and the magnitude ratio GMM, a method is proposed to obtain the optimal weights of the linear combination to enhance the localization ability. Additionally, because BRDPs contain information on direct paths and reflections, localizing a sound source in the azimuth, elevation and distance using the proposed GMBRDM is possible.

The remainder of this chapter is organized as follows. The next section discusses how the nonstationary sound source can influence the IPD and ILD. A simulation of a simplified environment is performed to verify the discussion in Section 2.3. Section 2.4 presents the formulation of the proposed GMBRDM. The summary is given in Section 2.5.

2.2 The Relation between the Nonstationary Sound Source and the BRDP

2.2.1 IPDs and ILDs of Stationary Sound Source

A linear time-invariant (LTI) room acoustic channel is represented by a K tapped finite impulse response (FIR) model $h(n)=\sum_{k=0}^{K-1} b_k \delta(n-k)$ as,

$$y(n)=\sum_{k=0}^{K-1} b_k x(n-k) \tag{2-1}$$

where $x(n)$ denotes sound signal emitted into the channel, $y(n)$ denotes the signal received by the ear, and b_k is the coefficients of the FIR model for the room impulse response (RIR) from sound source to an ear. Without lost of generality, the stationary input signal is assumed to be a complex exponential signal with frequency $\hat{\omega}$ and constant level A:

$$x(n)=Ae^{j\hat{\omega}n} \tag{2-2}$$

where $\hat{\omega}=\frac{2\pi\hat{k}}{N}$, which is the sampled frequency of an N-point STFT.

For such input, the corresponding output is

$$\begin{aligned} y(n) &= \sum_{k=0}^{K-1} b_k x(n-k) \\ &= \sum_{k=0}^{K-1} b_k A e^{j\hat{\omega}(n-k)} \\ &= \sum_{k=0}^{K-1} \left(b_k e^{-j\hat{\omega}k}\right) A e^{j\hat{\omega}n} \end{aligned} \tag{2-3}$$

Take the *N*-point STFT at frequency $\hat{\omega}$:

$$\begin{aligned} Y(\hat{\omega}) &= \sum_{n=0}^{N-1}\left(\sum_{k=0}^{K-1} b_k e^{-j\hat{\omega}k}\right) A e^{j\hat{\omega}n} e^{-j\hat{\omega}n} \\ &= NA\sum_{k=0}^{K-1} b_k e^{-j\hat{\omega}k} \end{aligned} \tag{2-4}$$

By denoting $y_L(n)$ and $y_R(n)$ as the signals received by left and right ears, respectively, and $Y_L(\hat{\omega})$ and $Y_R(\hat{\omega})$ are the STFT of $y_L(n)$ and $y_R(n)$, the IPD, $P(\hat{\omega})$, and ILD, $M(\hat{\omega})$, between $Y_L(\hat{\omega})$ and $Y_R(\hat{\omega})$ are

$$P(\hat{\omega}) = \angle\left(\frac{\sum_{k=0}^{K-1} b_{L,k} e^{-j\hat{\omega}k}}{\sum_{k=0}^{K-1} b_{R,k} e^{-j\hat{\omega}k}}\right)$$

and

$$M(\hat{\omega}) = \ln\left|\frac{\sum_{k=0}^{K-1} b_{L,k} e^{-j\hat{\omega}k}}{\sum_{k=0}^{K-1} b_{R,k} e^{-j\hat{\omega}k}}\right| \tag{2-5}$$

where $b_{L,k}$ and $b_{R,k}$ are the coefficients of FIR channel models, h_L and h_R, from the sound source to the left ear and the right ear, $h_L = \sum_{k=0}^{K-1} b_{L,k}\delta(n-k)$, $h_R = \sum_{k=0}^{K-1} b_{R,k}\delta(n-k)$ and $\angle(\cdot)$ denotes the phase value. Note that the operation of nature logarithm is taken for computing the magnitude ratio. As shown in (2-5), the IPD and ILD between $Y_L(\hat{\omega})$ and $Y_R(\hat{\omega})$ depend only on the frequency responses of the channels and the measured frequency, as discussed in related research.

2.2.2 IPDs and ILDs of Nonstationary Sound Source

Although the nonstationarity of a sound source can be tested in many different domains [50], this work only considers time domain variation. To model time domain variation of a sound source, the level of the complex exponential signal in (2-2) is assumed as time varying:

$$x(n) = A_n e^{j\phi} e^{j\hat{\omega} n} \tag{2-6}$$

where A_n is a time-varying sound level. Accordingly, the output $y(n)$ can be formulated as:

$$\begin{aligned} y(n) &= \sum_{k=0}^{K-1} b_k x(n-k) \\ &= \sum_{k=0}^{K-1} b_k A_{n-k} e^{j\phi} e^{j\hat{\omega}(n-k)} \\ &= \sum_{k=0}^{K-1} A_{n-k} b_k e^{-j\hat{\omega} k} e^{j\phi} e^{j\hat{\omega} n} \end{aligned} \tag{2-7}$$

Take the STFT at frequency $\hat{\omega}$:

$$\begin{aligned} Y(n,\hat{\omega}) &= \sum_{\tau=0}^{N-1} y(n+\tau) e^{-j\hat{\omega}(n+\tau)} \\ &= \sum_{\tau=0}^{N-1} \sum_{k=0}^{K-1} A_{n+\tau-k} b_k e^{-j\hat{\omega} k} e^{j\phi} e^{j\hat{\omega}(n+\tau)} e^{-j\hat{\omega}(n+\tau)} \\ &= \sum_{\tau=0}^{N-1} \sum_{k=0}^{K-1} A_{n+\tau-k} b_k e^{-j\hat{\omega} k} e^{j\phi} \end{aligned} \tag{2-8}$$

Hence, the ratio between $Y_L(n,\hat{\omega})$ and $Y_R(n,\hat{\omega})$ is

$$\frac{Y_L(n,\hat{\omega})}{Y_R(n,\hat{\omega})}=\frac{\sum_{\tau=0}^{N-1}\sum_{k=0}^{K-1}A_{n+\tau-k}b_{L,k}e^{-j\hat{\omega}k}}{\sum_{\tau=0}^{N-1}\sum_{k=0}^{K-1}A_{n+\tau-k}b_{R,k}e^{-j\hat{\omega}k}} \tag{2-9}$$

and $P(n,\hat{\omega})$ and $M(n,\hat{\omega})$ between $Y_L(n,\hat{\omega})$ and $Y_R(n,\hat{\omega})$ are

$$P(n,\hat{\omega})=\angle\left(\frac{\sum_{\tau=0}^{N-1}\sum_{k=0}^{K-1}A_{n+\tau-k}b_{L,k}e^{-j\hat{\omega}k}}{\sum_{\tau=0}^{N-1}\sum_{k=0}^{K-1}A_{n+\tau-k}b_{R,k}e^{-j\hat{\omega}k}}\right), \text{ and}$$

$$M(n,\hat{\omega})=\ln\left|\frac{\sum_{\tau=0}^{N-1}\sum_{k=0}^{K-1}A_{n+\tau-k}b_{L,k}e^{-j\hat{\omega}k}}{\sum_{\tau=0}^{N-1}\sum_{k=0}^{K-1}A_{n+\tau-k}b_{R,k}e^{-j\hat{\omega}k}}\right| \tag{2-10}$$

As shown in (2-10), the phase difference and magnitude ratio become content dependent when STFT is utilized and A_n is nonstationary.

2.2.3 Modeling the Nonstationary Sound Source Using Moving Pole Model

To analyze how nonstationarity of a sound source influences the IPD and ILD, a parameterized model for nonstationary sound is needed. Based on the discussion in [51] and [52], a nonstationary sound source in an analysis window can be expressed as a sum of moving pole models. In this work, the idea in [52] that approximate A_n as an exponent of polynomial is utilized. In [52],

$$A_n=e^{\sum_{t=0}^{N_a}a_t\left(\frac{n}{f_s}\right)^t} \tag{2-11}$$

where N_a is the degree of the polynomial, a_t is the coefficient of the polynomial and f_s denotes the sampling frequency. To simplify the analysis, we omit the terms of $t \geq 2$, as in [30]; hence, A_n is modeled as:

$$A_n = e^{a_0 + \frac{n}{f_s} a_1} \tag{2-12}$$

Substituting (2-12) into (2-8), $Y_L(n,\hat{\omega})$ can be rewritten as:

$$\begin{aligned}
Y_L(n,\hat{\omega}) = & \left[e^{a_0+\left(\frac{n}{f_s}\right)a_1} + e^{a_0+\left(\frac{n+1}{f_s}\right)a_1} + e^{a_0+\left(\frac{n+2}{f_s}\right)a_1} + \cdots + e^{a_0+\left(\frac{n+N-1}{f_s}\right)a_1} \right] b_{L,0} e^{-j\hat{\omega}0} e^{j\phi} \\
& + \left[e^{a_0+\left(\frac{n-1}{f_s}\right)a_1} + e^{a_0+\left(\frac{n}{f_s}\right)a_1} + e^{a_0+\left(\frac{n+1}{f_s}\right)a_1} + \cdots + e^{a_0+\left(\frac{n+N-2}{f_s}\right)a_1} \right] b_{L,1} e^{-j\hat{\omega}1} e^{j\phi} \\
& + \left[e^{a_0+\left(\frac{n-2}{f_s}\right)a_1} + e^{a_0+\left(\frac{n-1}{f_s}\right)a_1} + e^{a_0+\left(\frac{n}{f_s}\right)a_1} + \cdots + e^{a_0+\left(\frac{n+N-3}{f_s}\right)a_1} \right] b_{L,2} e^{-j\hat{\omega}2} e^{j\phi} \\
& \quad \vdots \\
& + \left[e^{a_0+\left(\frac{n-(K-1)}{f_s}\right)a_1} + e^{a_0+\left(\frac{n-(K-2)}{f_s}\right)a_1} + e^{a_0+\left(\frac{n-(K-3)}{f_s}\right)a_1} + \cdots + e^{a_0+\left(\frac{n+(N-K)}{f_s}\right)a_1} \right] b_{L,K-1} e^{-j\hat{\omega}(K-1)} e^{j\phi}
\end{aligned}$$

` (2-13)

This equation can be rearranged as:

$$\begin{aligned}
Y_L(n,\hat{\omega}) &= e^{a_0+\frac{n}{f_s}a_1}\left[1+e^{\frac{1}{f_s}a_1}++e^{\frac{2}{f_s}a_1}+\cdots+e^{\frac{(N-1)}{f_s}a_1}\right]b_{L,0}e^{-j\hat{\omega}0}e^{j\phi} \\
&\quad +e^{-\frac{a_1}{f_s}}e^{a_0+\frac{n}{f_s}a_1}\left[1+e^{\frac{1}{f_s}a_1}++e^{\frac{2}{f_s}a_1}+\cdots+e^{\frac{(N-1)}{f_s}a_1}\right]b_{L,1}e^{-j\hat{\omega}1}e^{j\phi} \\
&\quad +e^{-2\frac{a_1}{f_s}}e^{a_0+\frac{n}{f_s}a_1}\left[1+e^{\frac{1}{f_s}a_1}++e^{\frac{2}{f_s}a_1}+\cdots+e^{\frac{(N-1)}{f_s}a_1}\right]b_{L,2}e^{-j\hat{\omega}2}e^{j\phi} \\
&\qquad\qquad \vdots \\
&\quad +e^{-(K-1)\frac{a_1}{f_s}}e^{a_0+\frac{n}{f_s}a_1}\left[1+e^{\frac{1}{f_s}a_1}++e^{\frac{2}{f_s}a_1}+\cdots+e^{\frac{(N-1)}{f_s}a_1}\right]b_{L,K-1}e^{-j\hat{\omega}(K-1)}e^{j\phi} \\
&\quad =e^{a_0+\frac{n}{f_s}a_1}\frac{1-e^{N\frac{a_1}{f_s}}}{1-e^{\frac{a_1}{f_s}}}\left(\sum_{k=0}^{K-1}e^{-\frac{a_1}{f_s}k}b_{L,k}e^{-j\hat{\omega}k}\right)e^{j\phi}
\end{aligned} \tag{2-14}$$

Through the same procedure, we have

$$Y_R(n,\hat{\omega})=e^{a_0+\frac{n}{f_s}a_1}\frac{1-e^{N\frac{a_1}{f_s}}}{1-e^{\frac{a_1}{f_s}}}\left(\sum_{k=0}^{K-1}e^{-\frac{a_1}{f_s}k}b_{R,k}e^{-j\hat{\omega}k}\right)e^{j\phi} \tag{2-15}$$

and the ratio between $Y_L(n,\hat{\omega})$ and $Y_R(n,\hat{\omega})$ becomes:

$$\frac{Y_L(n,\hat{\omega})}{Y_R(n,\hat{\omega})}=\frac{\sum_{k=0}^{K-1}e^{-\frac{a_1}{f_s}k}b_{L,k}e^{-j\hat{\omega}k}}{\sum_{k=0}^{K-1}e^{-\frac{a_1}{f_s}k}b_{R,k}e^{-j\hat{\omega}k}} \tag{2-16}$$

Consequently, the IPD and ILD are

$$P(n,\hat{\omega})=\angle\left(\frac{\sum_{k=0}^{K-1}e^{-\frac{a_1}{f_s}k}b_{L,k}e^{-j\hat{\omega}k}}{\sum_{k=0}^{K-1}e^{-\frac{a_1}{f_s}k}b_{R,k}e^{-j\hat{\omega}k}}\right)\quad\text{and}$$

$$M(n,\hat{\omega})=\ln\left|\frac{\sum_{k=0}^{K-1}e^{-\frac{a_1}{f_s}k}b_{L,k}e^{-j\hat{\omega}k}}{\sum_{k=0}^{K-1}e^{-\frac{a_1}{f_s}k}b_{R,k}e^{-j\hat{\omega}k}}\right| \tag{2-17}$$

By observing (2-17), this study finds that the IPD and ILD values depend on the coefficient of the FIR models and the value of a_1, which is the slope of the nature logarithm of A_n.

2.3 Simulation Verification and Discussion of Proposed Model

2.3.1 Content Dependency of BRDPs Obtained from Nonstationary Sound Source

To verify the proposed analysis, a simplified simulation environment (Fig. 2-3) is assumed (Although the simplified environment is utilized as an example here, the following discussion of the relationship between BRDPs and nonstationary sound sources can be applied to general cases).

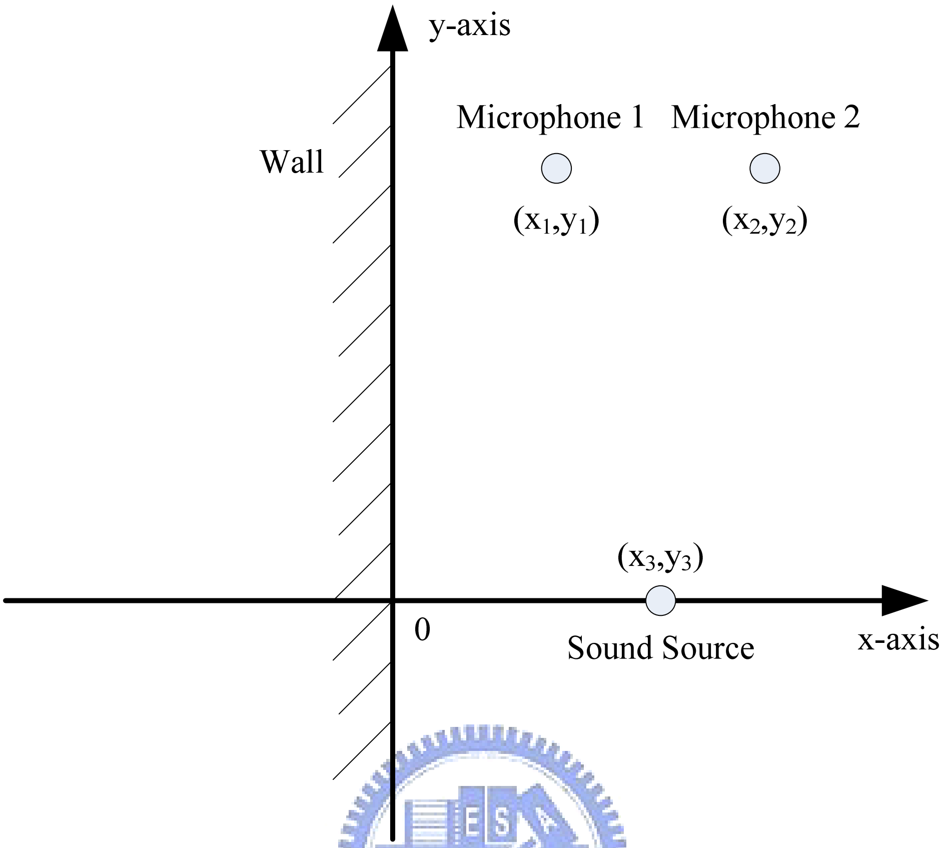


Figure 2-3 Simulation configuration.

As depicted in Fig. 2-3, the only cause of reflection is the only cause of reflection is the infinite wall located at $x = 0$. The two microphones are located at $(x_1, y_1, z_1) = (4.8\text{ m}, 0.5\text{ m}, 0\text{ m})$ and $(x_2, y_2, z_2) = (5.2\text{ m}, 0.5\text{ m}, 0\text{ m})$ and the sound source is located at $(x_3, y_3, z_3) = (5\text{ m}, 0\text{ m}, 0\text{ m})$. The models from the sound source to the microphones are simulated by the image method introduced in [47] with sound speed $c = 340\text{ m/s}$ and sampling rate $f_s = 8000\text{ Hz}$. The wall is assumed to be rigid. The simulated model is depicted in Fig. 2-4.

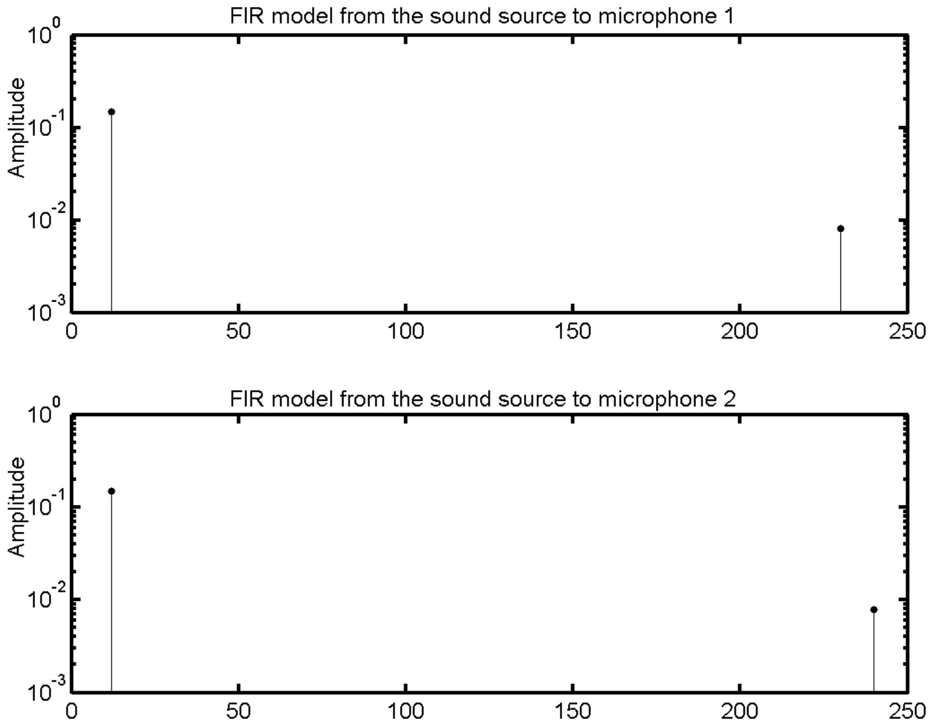
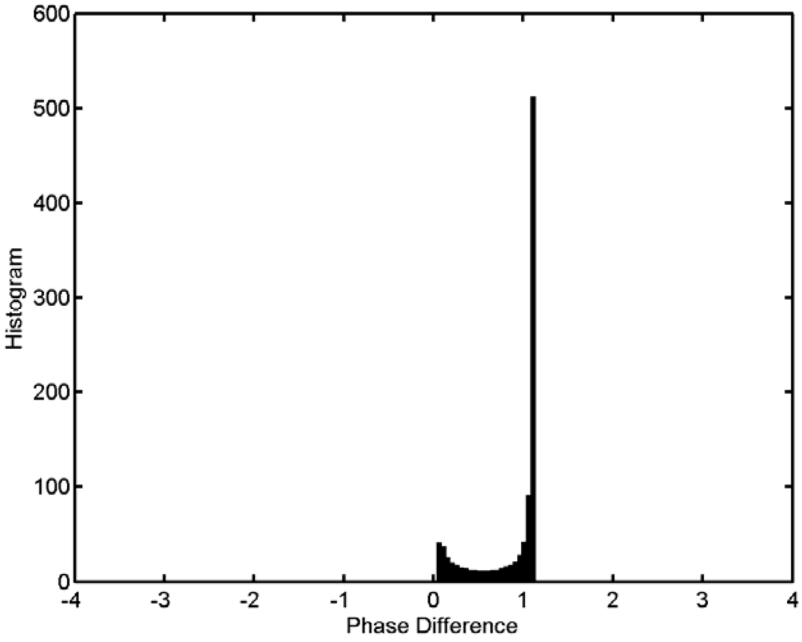
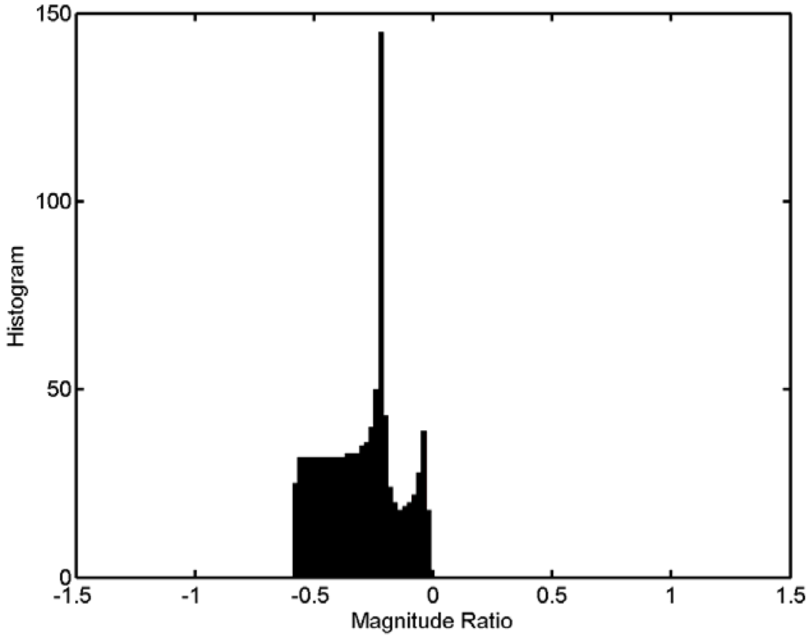


Figure 2-4 Simulated model from sound source to the microphones.

Two different sources are input into the simulation model to show the content dependency of IPD and ILD histograms. For the first source, the value of a_1 in the measured frames is uniformly distributed between $[-500,0]$. The IPDs and ILDs at a frequency of 140.625 Hz are computed 1000 times. Fig. 2-5 presents the histograms, which can represent the probability distribution, of IPDs and ILDs. The second source is similar to the first, except the value of a_1 is uniformly distributed between $[-500,200]$ in the measured frames. The histograms are illustrated in Fig. 2-6.

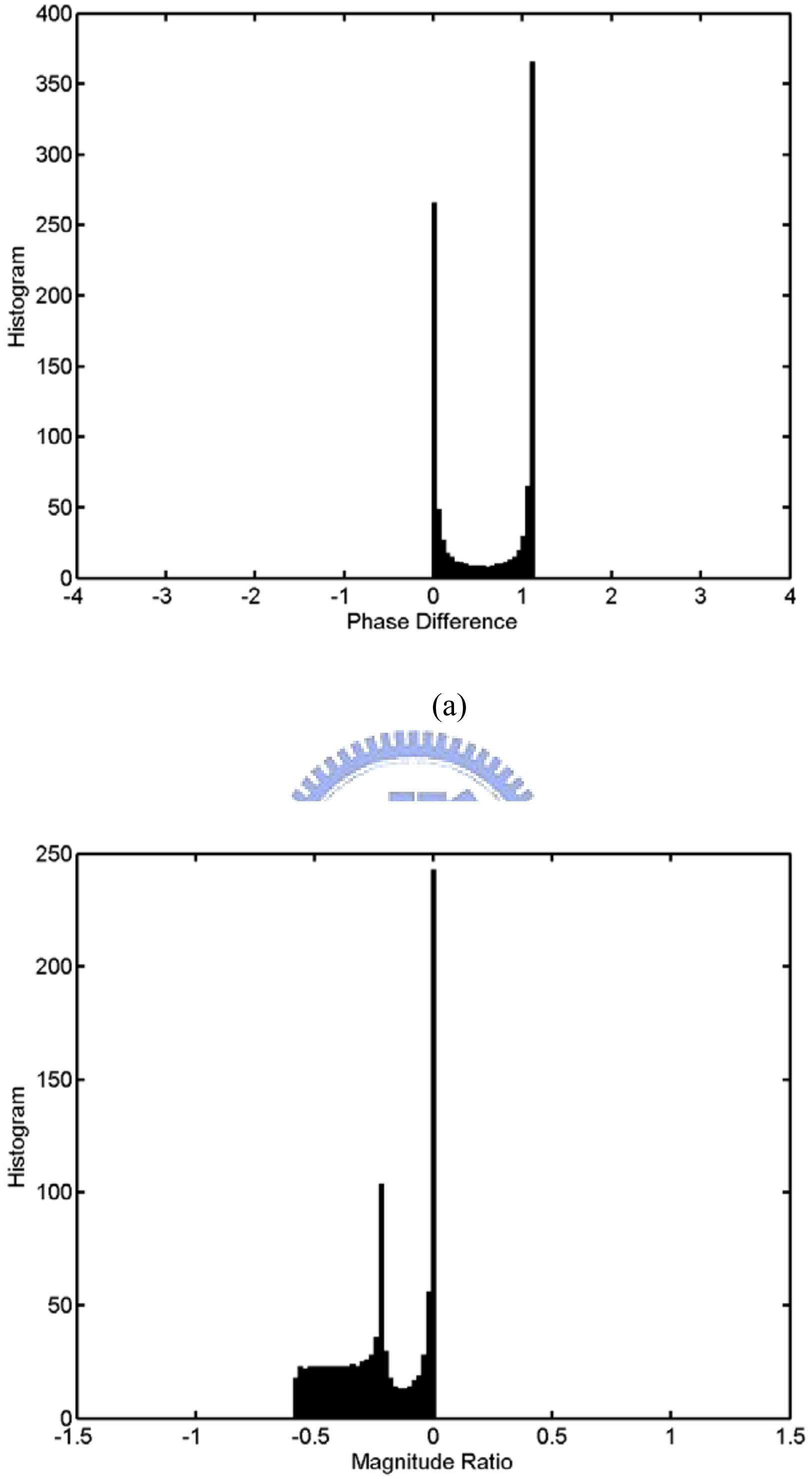


(a)



(b)

Figure 2-5 The histograms of IPDs and ILDs of the first sound source. (a) The histogram of IPDs of the first sound source. (b) The histogram of ILDs of the first sound source.



(a)

(b)

Figure 2-6 The histograms of IPDs and ILDs of the second sound source. (a) The histogram of IPDs of the second sound source. (b) The histogram of ILDs of the second sound source.

The simulation results in Figs. 2-5 and 2-6 demonstrate that when the sound sources are nonstationary, the IPD and ILD histograms depend on the content of the source signal. Therefore, conditions of the nonstationary sound source must be designed such that the BRDPs can be utilized for localization. In view of the aforementioned discussion, the sufficient condition is that the distribution of a_1 of the sound source must be stationary to make the sound source applicable for localization. Care must be exercised when using IPDs and ILDs obtained from nonstationary sound sources for sound source localization to avoid performance degradation.

2.3.2 The Formation of Peaks in the Distribution Patterns of IPDs

As shown by the simulation in Section 2.3.1, the distribution patterns of IPDs exhibit multiple peaks. This phenomenon also appears in the empirical results in real environment. The derivation result of (2-17) can be adopted to explain this phenomenon.

According to (2-17), there are several possible reasons to form peaks in the distribution patterns of IPDs. First, if a_1 of a sound source is concentrated at a certain value, a peak in the histogram will result. An obvious example is a stationary sound source. For a stationary sound source, $a_1 = 0$ for all measured frames, which makes IPD a fixed value, results in a peak in the distribution pattern.

Secondly, the term $e^{\frac{-a_1}{f_s}k}$ in (2-17) decreases as k increases when a_1 is positive. This means the weights of the reflection part in the channel model is reduced and the influence of the direct path are increased. Hence, when a_1 exceeds a certain level, the measured IPDs can be approximated as:

$$
\begin{aligned}
P(n,\hat{\omega}) &\approx \angle\left(\frac{e^{-\frac{a_1}{f_s}k_{D,1}}\, b_{L,k_{D,1}}\, e^{-j\hat{\omega}k_{D,1}}}{e^{-\frac{a_1}{f_s}k_{D,2}}\, b_{R,k_{D,2}}\, e^{-j\hat{\omega}k_{D,2}}}\right) \\
&= \angle\left(\frac{e^{-j\hat{\omega}k_{D,1}}}{e^{-j\hat{\omega}k_{D,2}}}\right)
\end{aligned}
\qquad (2\text{-}18)
$$

where $k_{D,1}$ and $k_{D,2}$ are propagation delay of the direct path from the sound source to microphones. Based on (2-18), the phase difference between direct paths from a sound source to microphones is emphasized and can dominate the measured IPDs. Since the IPDs are approximately the same for all a_1 exceed a certain level, a peak can be formed in the distribution pattern. This derivation explains why some previous research results of IPD-based time delay estimation suggested utilizing speech source onset to improve the accuracy [24]. On the contrary, when a_1 is negative, the value of $e^{\frac{-a_1}{f_s}k}$ increases with k. In this case, the influence of the direct path is suppressed and the reflections can dominate the measured IPDs.

The second simulation in Section 2.3.1 is utilized to interpret the relationship between a_1 and the IPD (Fig. 2-7).

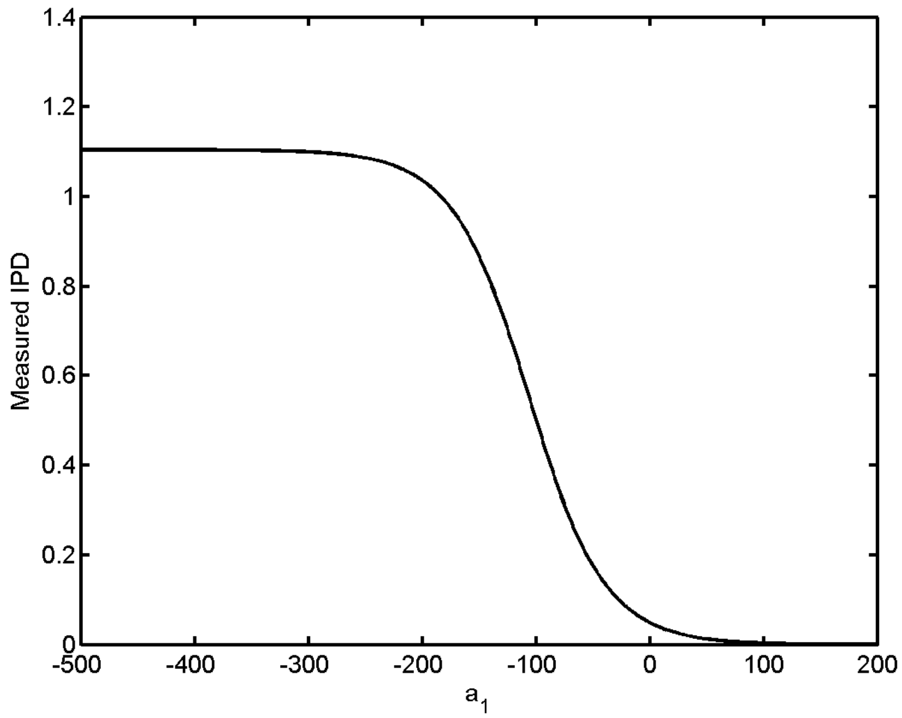


Figure 2-7 Relation between the value of a_1 and the IPD.

In Fig. 2-7, as $a_1 > 100$, the value of IPD approaches 0, which is the phase difference caused by the direct paths from the sound source to microphones. On the other hand, when $a_1 < -300$, the value converges to 1.1, representing the phase difference influenced by wall reflection. It is then easy to understand why there are two peaks at 0 and 1.1 in Fig. 2-6 (a). Generally, reflections appear later in the propagation model than direct paths, meaning that a negative value of a_1 is required to emphasize the effect of reflections. Consequently, the more the wall or boundary absorbs the energy of sound source, the smaller value of negative a_1 is required to emphasize the effect of reflections.

2.3.3 The Formation of Peaks in the Distribution Patterns of ILDs

Similar to the discussion of IPDs, the a_1 of a sound source is concentrated at a certain value, results in a peak in the ILD distribution pattern. However, ILDs behave quite different than IPDs when a_1 is either large or small. When a_1 is larger than a certain level, $M(n,\hat{\omega})$ can be approximated by

$$\begin{aligned} M(n,\hat{\omega}) &\approx \ln\left| \frac{e^{-\frac{a_1}{f_s}k_{D,1}} b_{L,k_{D,1}} e^{-j\hat{\omega}k_{D,1}}}{e^{-\frac{a_1}{f_s}k_{D,2}} b_{R,k_{D,2}} e^{-j\hat{\omega}k_{D,2}}} \right| \\ &= \ln\left| \frac{e^{-\frac{a_1}{f_s}k_{D,1}} b_{L,k_{D,1}}}{e^{-\frac{a_1}{f_s}k_{D,2}} b_{R,k_{D,2}}} \right| \\ &= \frac{-a_1(k_{D,1}-k_{D,2})}{f_s} + \ln\frac{b_{L,k_{D,1}}}{b_{R,k_{D,2}}} \end{aligned} \qquad (2\text{-}19)$$

Therefore, the relationship between ILDs and a_1 is approximately linear (with a slope of $\frac{-(k_{D,1}-k_{D,2})}{f_s}$) when a_1 is larger than a certain level. Hence, if the slope is 0 (meaning $k_{D,1}=k_{D,2}$), it will cause a peak in the ILD histogram. Similar to IPDs, when a_1 is smaller than a certain level, the influence of the direct path is de-emphasized and the reflection part starts dominating the measured ILDs. The second simulation in Section 2.3.1 is again utilized as an example of the discussion above. Figure 2-8 shows the simulation results for the relationship between the value of a_1 and the ILD.

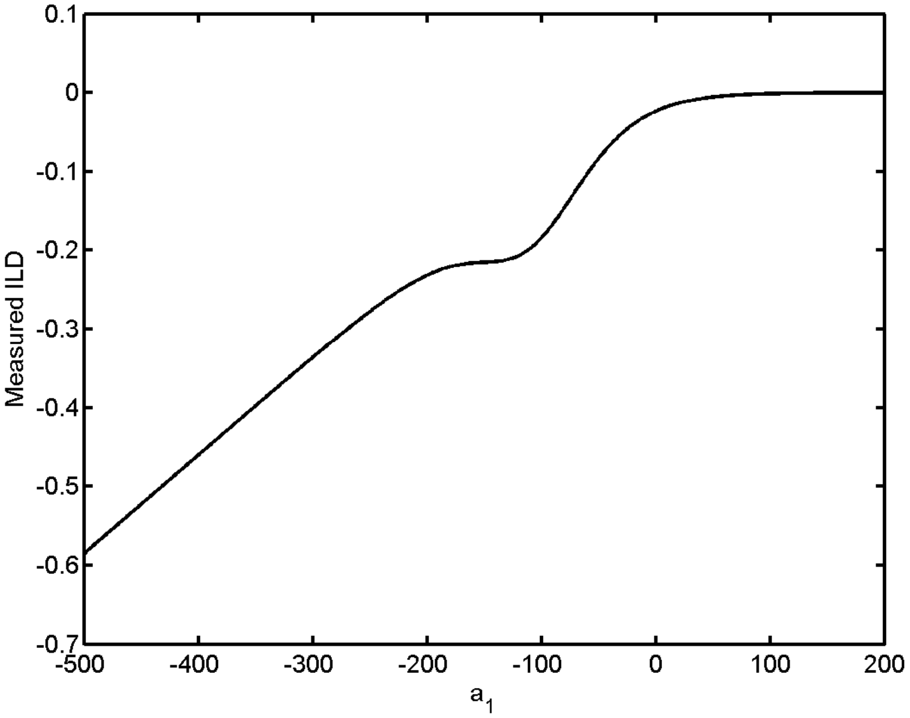


Figure 2-8 Relation between the value of a_1 and the ILD.

In Fig. 2-8, when $a_1 > 100$, the measured ILD is about 0 because the simulation sets $k_{D,1} - k_{D,2} = 0$ and $b_{L,k_{D,1}} = b_{L,k_{D,2}}$. This results in a peak at 0 in the histogram, as shown in Fig. 2-6 (b). In addition, when $a_1 < -300$, the measured ILDs change linearly with the value of a_1, resulting in a flat area in Fig. 2-6 (b).

2.3.4 Localization of Nonstationary Sound Source Using BRDPs

As mentioned in Chapter 1, detecting the location of sound sources presented at median plane or on a “cone of confusion” is difficult when only IPDs and ILDs of direct paths are utilized. However, sound sources at different locations can propagate through different reflections and with the property of nonstationary sound source discussed above, the nonstationary sound can result in distinguishable distribution patterns. Consequently, it is possible to detect the location of the sound sources in the azimuth, elevation, and distance using BRDPs.

2.4 GMBRDM for Nonstationary Sound Source Localization

As discussed in Section 2.3.1, if the environment and head position are unchanged and the distribution of a_1 of the sound source is stationary, using BRDPs for sound source localization is possible. Sections 2.3.2 and 2.3.3 also show that BRDPs can be non-Gaussian and contain multiple peaks. Consequently, modeling these distribution patterns as a simple distribution pattern (such as a single Gaussian distribution) can eliminate important details. Utilizing a high-resolution normalized histogram to model the distribution pattern requires considerable computational of memory. In this work, GMMs are employed to model BRDPs (called the GMBRDM) to reduce the memory requirement through parameterization.

2.4.1 The Training Procedure of the Proposed GMBRDM

Let $P_x(n_f, \omega_b)$ and $M_x(n_f, \omega_b)$ denote the phase difference and magnitude ratio obtained at frame n_f respectively for constructing GMM at frequency ω_b, $b \in \{1,...,B\}$, which means B frequencies are utilized to construct the model. The phase difference and magnitude ratio GMMs are defined as the weighted sum of N_1 and N_2 mixtures of Gaussian component densities:

$$G(\boldsymbol{P}_x(n_f) \mid \boldsymbol{\lambda}_P) = \sum_{i=1}^{N_1} \rho_{P,i} g_i(\boldsymbol{P}_x(n_f)) \tag{2-20}$$

$$G(\boldsymbol{M}_x(n_f) \mid \boldsymbol{\lambda}_M) = \sum_{i=1}^{N_2} \rho_{M,i} g_i(\boldsymbol{M}_x(n_f)) \tag{2-21}$$

where $\boldsymbol{P}_x(n_f) = [P_x(n_f, \omega_1) \;\; \cdots \;\; P_x(n_f, \omega_B)]^T$,

$\boldsymbol{M}_x(n_f) = [M_x(n_f, \omega_1) \;\; \cdots \;\; M_x(n_f, \omega_B)]^T$. $\rho_{P,i}$ and $\rho_{M,i}$ are the weights of i^{th} mixture, and $g_i(\boldsymbol{P}_x(n_f))$ and $g_i(\boldsymbol{M}_x(n_f))$ are the Gaussian density functions. Notably, the mixture weights must satisfy the constraints:

$$\sum_{i=1}^{N_1} \rho_{P,i} = 1 \text{ and } \sum_{i=1}^{N_2} \rho_{M,i} = 1 \tag{2-22}$$

The terms $\boldsymbol{\lambda}_P$ and $\boldsymbol{\lambda}_M$ represent the parameters of N_1 and N_2 component densities.

$$\boldsymbol{\lambda}_P = \{\boldsymbol{\rho}_P, \boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P\} \text{ and } \boldsymbol{\lambda}_M = \{\boldsymbol{\rho}_M, \boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M\} \tag{2-23}$$

where

$\boldsymbol{\rho}_P = [\rho_{P,1} \;\; \cdots \;\; \rho_{P,N_1}]$ denotes the phase difference mixture weight vector with dimensions $1 \times N_1$.

$\boldsymbol{\rho}_M = [\rho_{M,1} \;\; \cdots \;\; \rho_{M,N_2}]$ denotes the magnitude ratio mixture weight vector with dimensions $1 \times N_2$.

$\boldsymbol{\mu}_P = [\boldsymbol{\mu}_{P,1} \;\; \cdots \;\; \boldsymbol{\mu}_{P,N_1}]$ denotes the phase difference mean matrix with dimensions $B \times N_1$.

$\boldsymbol{\mu}_M = [\boldsymbol{\mu}_{M,1} \;\; \cdots \;\; \boldsymbol{\mu}_{M,N_2}]$ denotes the magnitude ratio mean matrix with dimensions $B \times N_2$.

$\mathbf{\Sigma}_P = \left[\mathbf{\Sigma}_{P,1} \quad \cdots \quad \mathbf{\Sigma}_{P,N_1}\right]$ denotes the phase difference covariance matrix with dimensions $B \times BN_1$.

$\mathbf{\Sigma}_M = \left[\mathbf{\Sigma}_{M,1} \quad \cdots \quad \mathbf{\Sigma}_{M,N_2}\right]$ denotes the magnitude ratio covariance matrix with dimensions $B \times BN_2$.

The parameters $\boldsymbol{\lambda}_P$ and $\boldsymbol{\lambda}_M$ in (2-23) can be estimated by the iterative EM algorithm [53] which guarantees a monotonic increase in the model's log-likelihood value. By denoting the training sequence length as N_F, the iterative procedure can be divided into the expectation step and maximum step:

Expectation step:

$$G(i \mid \boldsymbol{P}_x(n_f), \boldsymbol{\lambda}_P) = \rho_{P,i} g_i(\boldsymbol{P}_x(n_f)) \Big/ \sum_{i=1}^{N_1} \rho_{P,i} g_i(\boldsymbol{P}_x(n_f)) \tag{2-24}$$

$$G(i \mid \boldsymbol{M}_x(n_f), \boldsymbol{\lambda}_M) = \rho_{M,i} g_i(\boldsymbol{M}_x(n_f)) \Big/ \sum_{i=1}^{N_2} \rho_{M,i} g_i(\boldsymbol{M}_x(n_f)) \tag{2-25}$$

where $G(i \mid \boldsymbol{P}_x(n_f), \boldsymbol{\lambda}_P)$ and $G(i \mid \boldsymbol{M}_x(n_f), \boldsymbol{\lambda}_M)$ are *a posteriori* probabilities.

Maximization step:

(i). Estimate the mixture weights:

$$\rho_{P,i} = 1/N_F \sum_{n_f=1}^{N_F} G(i \mid \boldsymbol{P}_x(n_f), \boldsymbol{\lambda}_P) \tag{2-26}$$

$$\rho_{M,i} = 1/N_F \sum_{n_f=1}^{N_F} G\left(i \mid \boldsymbol{M}_x\left(n_f\right), \boldsymbol{\lambda}_M\right) \tag{2-27}$$

(ii). Estimate the mean vector:

$$\boldsymbol{\mu}_{P,i} = \sum_{n_f=1}^{N_F} G\left(i \mid \boldsymbol{P}_x\left(n_f\right), \boldsymbol{\lambda}_P\right) \boldsymbol{P}_x\left(n_f\right) \Big/ \sum_{n_f=1}^{n_F} G\left(i \mid \boldsymbol{P}_x\left(n_f\right), \boldsymbol{\lambda}_P\right) \tag{2-28}$$

$$\boldsymbol{\mu}_{M,i} = \sum_{n_f=1}^{N_F} G\left(i \mid \boldsymbol{M}_x\left(n_f\right), \boldsymbol{\lambda}_M\right) \boldsymbol{M}_x\left(n_f\right) \Big/ \sum_{n_f=1}^{N_F} G\left(i \mid \boldsymbol{M}_x\left(n_f\right), \boldsymbol{\lambda}_M\right) \tag{2-29}$$

(iii). Estimate the variances:

$$\sigma_{P,i}^2(\omega_b) = \left(\sum_{n_f=1}^{N_F} G\left(i \mid \boldsymbol{P}_x\left(n_f\right), \boldsymbol{\lambda}_P\right) P_x^{\,2}\left(n_f, \omega_b\right) \Big/ \sum_{n_f=1}^{N_F} G\left(i \mid \boldsymbol{P}_x\left(n_f\right), \boldsymbol{\lambda}_P\right)\right) - \mu_{P,i}^{\;2}(\omega_b) \tag{2-30}$$

$$\sigma_{M,i}^2(\omega_b) = \left(\sum_{n_f=1}^{N_F} G\left(i \mid \boldsymbol{M}_x\left(n_f\right), \boldsymbol{\lambda}_M\right) M_x^{\,2}\left(n_f, \omega_b\right) \Big/ \sum_{n_f=1}^{N_F} G\left(i \mid \boldsymbol{M}_x\left(n_f\right), \boldsymbol{\lambda}_M\right)\right) - \mu_{M,i}^{\;2}(\omega_b) \tag{2-31}$$

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 [54]. This work utilizes an accelerated K-means algorithm proposed by Elkan [55], which can significantly reduce the computational power requirement.

The proposed GMBRDM at location l is defined as the linear combination of the phase difference GMM and the magnitude ratio GMM obtained at location l:

$$GMBRDM(l) = \alpha_P G\left(\boldsymbol{P}_x(n_f) \mid \boldsymbol{\lambda}_P(l)\right) + \alpha_M G\left(\boldsymbol{M}_x(n_f) \mid \boldsymbol{\lambda}_M(l)\right) \qquad (2\text{-}32)$$

where α_P and α_M represent the weighting factors. The values of α_P and α_M 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_P and α_M are determined by the following formula:

$$\min\left\{\sum_{\boldsymbol{q}_P} \alpha_P \left\{\mathbf{C}_P(\boldsymbol{q}_P)\mathbf{U}\mathbf{C}_P(\boldsymbol{q}_P)^T\right\} + \sum_{\boldsymbol{q}_M} \alpha_M \left\{\mathbf{C}_M(\boldsymbol{q}_M)\mathbf{U}\mathbf{C}_M(\boldsymbol{q}_M)^T\right\}\right\}$$

$$s.t.\ \alpha_P \times \alpha_M = 1, \alpha_P > 0, \alpha_M > 0 \qquad (2\text{-}33)$$

where $\boldsymbol{q}_P \in Q_P$ and $\boldsymbol{q}_M \in Q_M$ are the B dimensional random vectors in the operation ranges, Q_P and Q_M.

$\mathbf{C}_P(\boldsymbol{q}_P) = [C(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(1)) \quad C(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(2)) \quad \cdots \quad C(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(L))]$,

$\mathbf{C}_M(\boldsymbol{q}_M) = [C(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(1)) \quad C(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(2)) \quad \cdots \quad C(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(L))]$, and

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

In addition,

$$C(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(l)) = H(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(l)) \Big/ \sqrt{\sum_{\boldsymbol{q}_P} H^2(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(l))} \tag{2-34}$$

$$C(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(l)) = H(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(l)) \Big/ \sqrt{\sum_{\boldsymbol{q}_M} H^2(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(l))} \tag{2-35}$$

$$H(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(l)) = G(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(l)) - \left(\sum_{\boldsymbol{q}_P} G(\boldsymbol{q}_P \mid \boldsymbol{\lambda}_P(l)) \Big/ N(\boldsymbol{q}_P) \right) \text{, and}$$

$$H(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(l)) = G(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(l)) - \left(\sum_{\boldsymbol{q}_M} G(\boldsymbol{q}_M \mid \boldsymbol{\lambda}_M(l)) \Big/ N(\boldsymbol{q}_M) \right)$$

(2-36)

where $N(\boldsymbol{q}_P)$ and $N(\boldsymbol{q}_M)$ denote the total selected numbers of $\boldsymbol{q}_P$ and $\boldsymbol{q}_M$.

The values of α_P and α_M can be obtained by solving (2-33) as:

$$\alpha_P = \sqrt{\sum_{\boldsymbol{q}_M} \mathbf{C}_M(\boldsymbol{q}_M)\mathbf{U}\mathbf{C}_M(\boldsymbol{q}_M)^T \Big/ \sum_{\boldsymbol{q}_P} \mathbf{C}_P(\boldsymbol{q}_P)\mathbf{U}\mathbf{C}_P(\boldsymbol{q}_P)^T} \tag{2-37}$$

$$\alpha_M = \sqrt{\sum_{\boldsymbol{q}_P} \mathbf{C}_P(\boldsymbol{q}_P)\mathbf{U}\mathbf{C}_P(\boldsymbol{q}_P)^T \Big/ \sum_{\boldsymbol{q}_M} \mathbf{C}_M(\boldsymbol{q}_M)\mathbf{U}\mathbf{C}_M(\boldsymbol{q}_M)^T} \tag{2-38}$$

The proofs of (2-37) and (2-38) are shown in the appendix.

2.4.2 The Testing Procedure of the Proposed GMBRDM

The location is determined by finding the maximum *a posteriori* location probability for a given observation sequence:

$$\hat{l} = \arg\max_{1\le l\le L} GMBRDM(l) = \arg\max_{1\le l\le L} \alpha_P G(\boldsymbol{\lambda}_P(l) \mid \mathbf{P}_Y) + \alpha_M G(\boldsymbol{\lambda}_M(l) \mid \mathbf{M}_Y)$$

$$= \arg\max_{1\le l\le L} \alpha_P \left(G(\mathbf{P}_Y \mid \boldsymbol{\lambda}_P(l)) p(\boldsymbol{\lambda}_P(l)) / p(\mathbf{P}_Y)\right) + \alpha_M \left(G(\mathbf{M}_Y \mid \boldsymbol{\lambda}_M(l)) p(\boldsymbol{\lambda}_M(l)) / p(\mathbf{M}_Y)\right)$$

(2-39)

where $\mathbf{P}_Y = \{\boldsymbol{P}_Y(1), \cdots, \boldsymbol{P}_Y(N_V)\}$ and $\mathbf{M}_Y = \{\boldsymbol{M}_Y(1), \cdots, \boldsymbol{M}_Y(N_V)\}$ are the phase difference and magnitude ratio computed from the testing sequences denoted as $Y_1(\omega)$ and $Y_2(\omega)$, and N_V denotes the testing sequence length. The probabilities $p(\boldsymbol{\lambda}_P(l))$ and $p(\boldsymbol{\lambda}_M(l))$ can be selected as $1/L$ since the probability in each location is equally likely for a blind search. Moreover, because the probability densities $p(\mathbf{P}_Y)$ and $p(\mathbf{M}_Y)$ are the same for all location models, the detection rule can be recast as:

$$\hat{l} = \arg\max_{1\le l\le L} \alpha_P \prod_{n_v=1}^{N_V} G(\boldsymbol{P}_Y(n_v) \mid \boldsymbol{\lambda}_P(l)) + \alpha_M \prod_{n_v=1}^{N_V} G(\boldsymbol{M}_Y(n_v) \mid \boldsymbol{\lambda}_M(l)) \qquad \text{(2-40)}$$

2.5 Summary

In this chapter, the relation between the nonstationary sound source and the BRDPs are discussed. Moreover, based on the discussion, a model, named GMBRDM is proposed for nonstationary sound source localization. Theoretically, the GMBRDM is capable of localizing sound source in azimuth, elevation, and distance. The performance of the proposed method is examined in Chapter 4.

Appendix

Proofs of (2-37) and (2-38):

The problem is formulated as

$$\min\left\{\sum_{\boldsymbol{q}_p}\alpha_p\left\{\mathbf{C}_P(\boldsymbol{q}_P)\mathbf{U}\mathbf{C}_P(\boldsymbol{q}_P)^T\right\}+\sum_{\boldsymbol{q}_M}\alpha_M\left\{\mathbf{C}_M(\boldsymbol{q}_M)\mathbf{U}\mathbf{C}_M(\boldsymbol{q}_M)^T\right\}\right\}$$

$$s.t.\ \alpha_P\alpha_M=1,\alpha_P>0,\alpha_M>0 \qquad \text{(A-1)}$$

According to the constraint, set $\alpha_M = 1/\alpha_P$. Then, the cost function becomes,

$$\min\left\{\sum_{\boldsymbol{q}_p}\alpha_p\left\{\mathbf{C}_P(\boldsymbol{q}_P)\mathbf{U}\mathbf{C}_P(\boldsymbol{q}_P)^T\right\}+\sum_{\boldsymbol{q}_M}\frac{1}{\alpha_p}\left\{\mathbf{C}_M(\boldsymbol{q}_M)\mathbf{U}\mathbf{C}_M(\boldsymbol{q}_M)^T\right\}\right\} \qquad \text{(A-2)}$$

Setting the first derivative with respect to α_P be zero gives

$$\sum_{\boldsymbol{q}_p}\mathbf{C}_P(\boldsymbol{q}_P)\mathbf{U}\mathbf{C}_P(\boldsymbol{q}_P)^T-\sum_{\boldsymbol{q}_M}\alpha_p{}^{-2}\mathbf{C}_M(\boldsymbol{q}_M)\mathbf{U}\mathbf{C}_M(\boldsymbol{q}_M)^T=0 \qquad \text{(A-3)}$$

Therefore,

$$\alpha_P=\sqrt{\frac{\sum_{\boldsymbol{q}_M}\mathbf{C}_M(\boldsymbol{q}_M)\mathbf{U}\mathbf{C}_M(\boldsymbol{q}_M)^T}{\sum_{\boldsymbol{q}_p}\mathbf{C}_P(\boldsymbol{q}_P)\mathbf{U}\mathbf{C}_P(\boldsymbol{q}_P)^T}} \qquad \text{(A-4)}$$

$$\alpha_M = \frac{1}{\alpha_p} = \sqrt{\frac{\sum_{\boldsymbol{q}_p} \mathbf{C}_P\left(\boldsymbol{q}_P\right)\mathbf{U}\mathbf{C}_P\left(\boldsymbol{q}_P\right)^T}{\sum_{\boldsymbol{q}_M} \mathbf{C}_M\left(\boldsymbol{q}_M\right)\mathbf{U}\mathbf{C}_M\left(\boldsymbol{q}_M\right)^T}} \qquad \text{(A-5)}$$

Chapter 3

Indoor Sound Field Feature Matching for Robot's Location and Orientation Detection

3.1 Introduction

Indoor robot localization is an important issue in the field of robotics. Various equipments, such as camera, radio frequency identification (RFID), infrared (IR), ultra sonic sensor, laser, wireless LAN based methods and inertial navigation sensor have been adopted to provide different solutions [57 - 64].

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 [65, 66]. The BRDPs introduced in Chapter 2 are treated as sound field features and this chapter. Therefore, this chapter 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.

3.1.1 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 [10, 11, 67, 68] were discussed for robot localization application [65]. In general, sound-based robot localization system uses a loudspeaker 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 [65, 66]. 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, as discussed in Chapter 2. 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.

3.1.2 The Proposed Method

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. Instead of trying to eliminate the influence of reflection and diffraction, this work treats the distribution patterns of phase difference and magnitude ratio as a local feature and uses it to detect the robot's location and orientation. As discussed in Chapter 2, the complex propagation behaviors of a sound source result in location or orientation dependent phase difference and magnitude ratio distributions. This work adopts GMMs to model these 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 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 chapter is organized as follows. The Section 3.2 introduces the overall system architecture. Section 3.3 describes the design of the directional sound pattern for orientation detection. Section 3.4 presents the formulations of the proposed RLM and ROM. Finally, a summary is drawn in Section 3.5.

3.2 System Architecture

As shown in Fig. 3-1, the proposed system contains two loudspeakers 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. The sound patterns generated by Speaker 1 (SP1) are received by the RLODA and the RLMs can be obtained by location dependent phase difference and magnitude ratio distributions 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 3.3. 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.

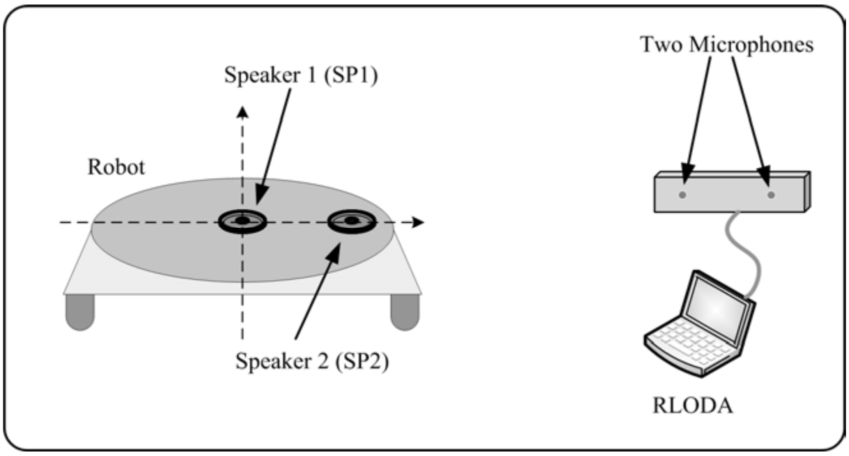


Figure 3-1 Speaker and microphone configuration of the proposed system.

Figure 3-2 depicts the overall system architecture. Stage I in Fig. 3-2 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 loudspeakers to obtain a pre-recorded database. Since the sound is recorded by the two microphones, the information of the sound field features and microphone response can be obtained by this database.

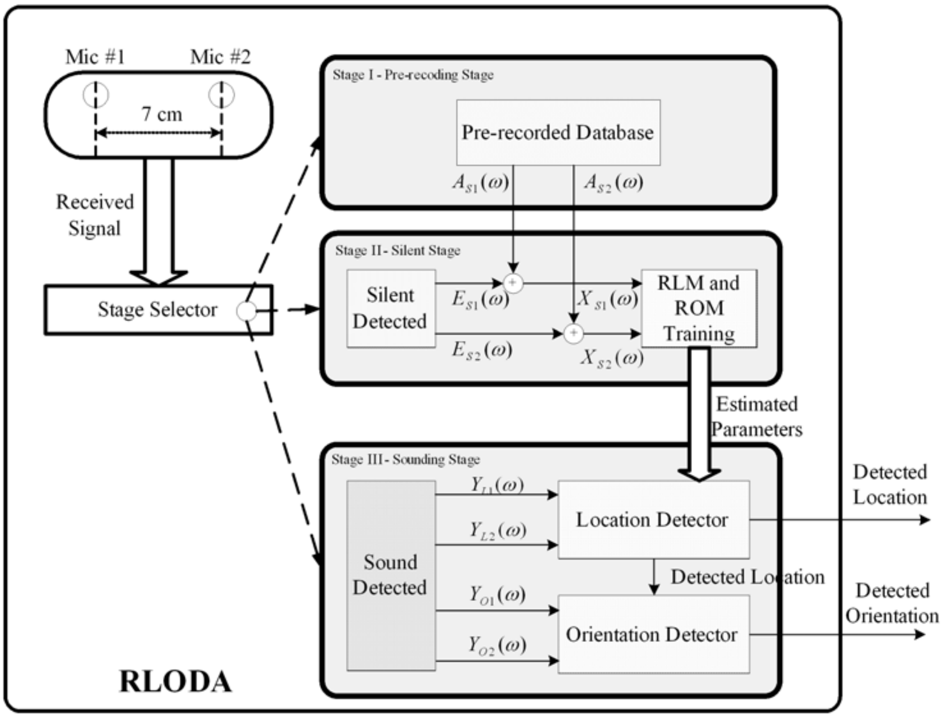


Figure 3-2 Overall system architecture.

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-recorded 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. Figure 3-3 illustrates the flowchart of proposed system.

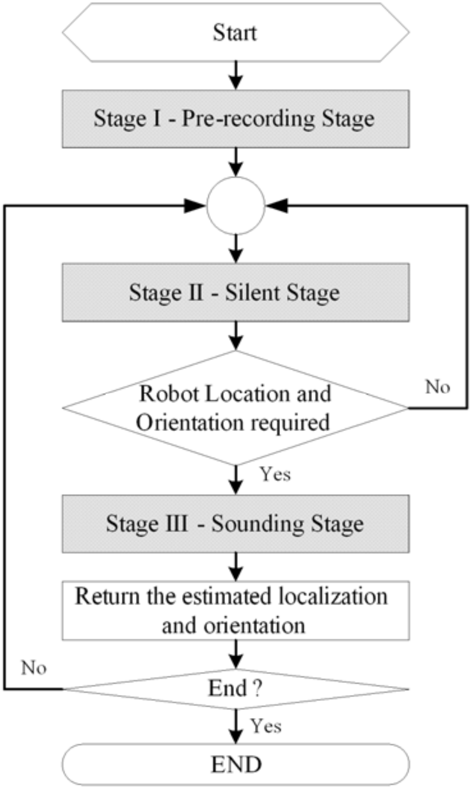

Start
Stage I - Pre-recording Stage
Stage II - Silent Stage
Robot Location and Orientation required
No
Yes
Stage III - Sounding Stage
Return the estimated localization and orientation
No
End ?
Yes
END

Figure 3-3 Flowchart of the proposed system.

3.3 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 [69, 70] 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. Figure 3-4 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. 3-4, 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.

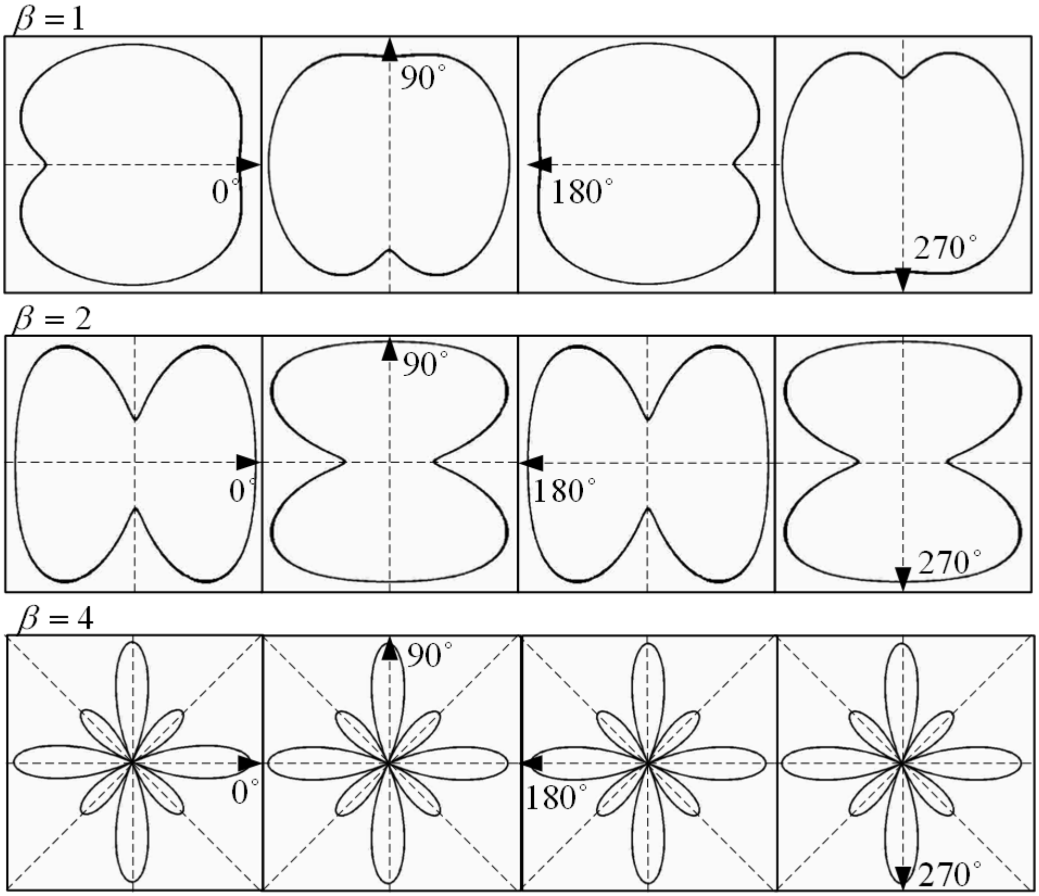


Figure 3-4 Relations between β and the sound pattern.

Therefore, the generated sound can only be symmetrical along one axis ($\beta = 1$) to avoid confusion in orientation detection. Consequently, this work proposes a method that utilizes two loudspeakers to generate the sound pattern that conforms to the constraint by:

$$\begin{aligned} J_{SP1}(n) &= J(n) \\ J_{SP2}(n) &= 0.5 \times J(n) \end{aligned} \tag{3-1}$$

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

Figure 3-5 depicts the simulation of the generated sound pattern of the proposed system based on the sound propagation theories in [71] 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. 3-5, the generated sound pattern is symmetric along only one axis and is suitable for robot's orientation detection.

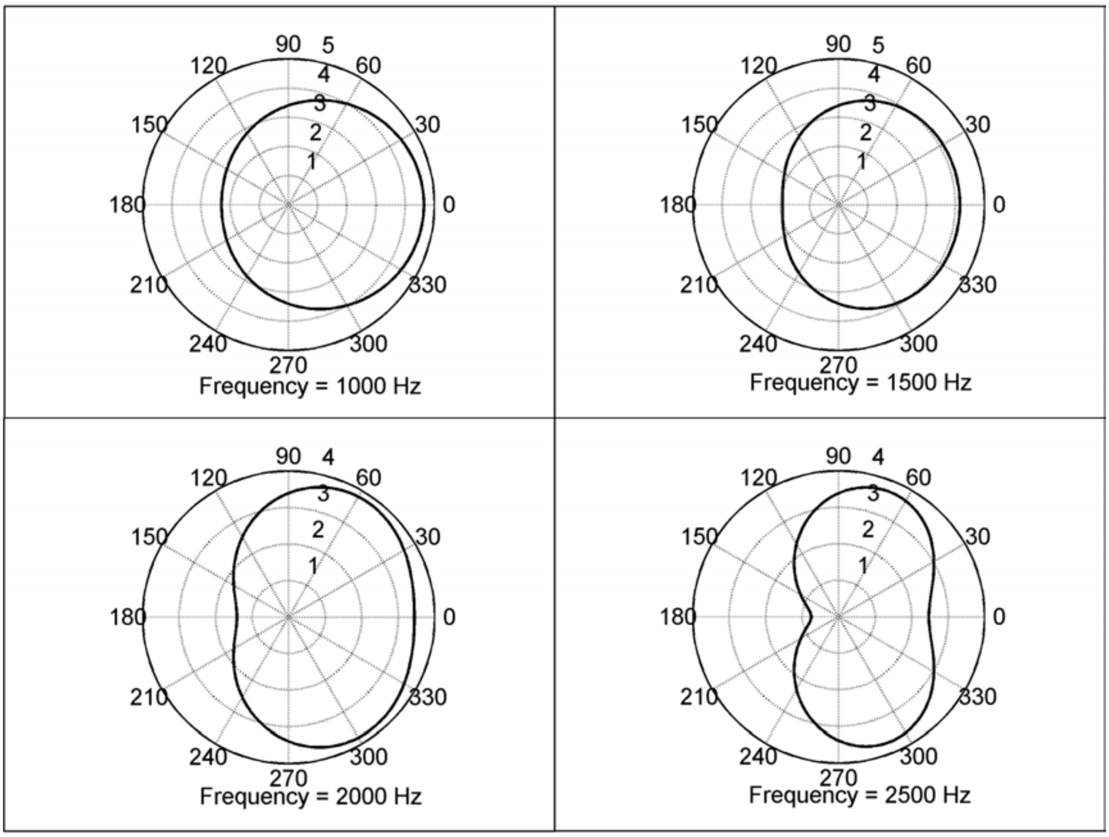


Figure 3-5 Simulation of generated sound pattern.

3.4 Robot Localization Model (RLM) and Robot Orientation Model (ROM)

3.4.1 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. $P_{Sx}(n_f,\omega_b)$ and $M_{Sx}(n_f,\omega_b)$ denote the phase difference and magnitude ratio at frame n_f respectively for constructing RLM ($S=L$) or ROM ($S=O$) at frequency ω_b, $b\in\{1,\dots,B\}$. The GMMs are defined as the weighted sum of N_1 and N_2 mixtures of Gaussian component densities shown below,

$$G(\boldsymbol{P}_{Sx}(n_f)\,|\,\boldsymbol{\lambda}_{SP})=\sum_{i=1}^{N_1}\rho_{SP,i}\,g_i(\boldsymbol{P}_{Sx}(n_f)) \tag{3-2}$$

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

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

$$\sum_{i=1}^{N_1}\rho_{SP,i}=1 \ \text{ and } \ \sum_{i=1}^{N_2}\rho_{SM,i}=1 \tag{3-4}$$

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

$$\boldsymbol{\lambda}_{SP}=\{\boldsymbol{\rho}_{SP},\boldsymbol{\mu}_{SP},\boldsymbol{\Sigma}_{SP}\} \text{ and } \boldsymbol{\lambda}_{SM}=\{\boldsymbol{\rho}_{SM},\boldsymbol{\mu}_{SM},\boldsymbol{\Sigma}_{SM}\} \tag{3-5}$$

where

$\boldsymbol{\rho}_{SP}=\lfloor \rho_{SP,1} \quad \cdots \quad \rho_{SP,N_1} \rfloor$ denotes the phase difference mixture weight vector with dimensions $1\times N_1$.

$\boldsymbol{\rho}_{SM}=\lfloor \rho_{SM,1} \quad \cdots \quad \rho_{SM,N_2} \rfloor$ denotes the magnitude ratio mixture weight vector with dimensions $1\times N_2$.

$\boldsymbol{\mu}_{SP}=\lfloor \boldsymbol{\mu}_{SP,1} \quad \cdots \quad \boldsymbol{\mu}_{SP,N_1} \rfloor$ denotes the phase difference mean matrix with dimensions $B\times N_1$.

$\boldsymbol{\mu}_{SM}=\lfloor \boldsymbol{\mu}_{SM,1} \quad \cdots \quad \boldsymbol{\mu}_{SM,N_2} \rfloor$ denotes the magnitude ratio mean matrix with dimensions $B\times N_2$.

$\boldsymbol{\Sigma}_{SP}=\lfloor \boldsymbol{\Sigma}_{SP,1} \quad \cdots \quad \boldsymbol{\Sigma}_{SP,N_1} \rfloor$ denotes the phase difference covariance matrix with dimensions $B\times BN_1$

$\boldsymbol{\Sigma}_{SM}=\lfloor \boldsymbol{\Sigma}_{SM,1} \quad \cdots \quad \boldsymbol{\Sigma}_{SM,N_2} \rfloor$ denotes the magnitude ratio covariance matrix with dimensions $B\times BN_2$.

The parameters $\boldsymbol{\lambda}_{SP}$ and $\boldsymbol{\lambda}_{SM}$ in (3-5) can be estimated by the iterative EM algorithm, 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:

$$G\left(i \mid \boldsymbol{P}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SP}\right)=\rho_{SP,i}\,g_i\left(\boldsymbol{P}_{Sx}\left(n_f\right)\right)\Big/\sum_{i=1}^{N_1}\rho_{SP,i}\,g_i\left(\boldsymbol{P}_{Sx}\left(n_f\right)\right) \tag{3-6}$$

$$G\left(i \mid \boldsymbol{M}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SM}\right)=\rho_{SM,i}\,g_i\left(\boldsymbol{M}_{Sx}\left(n_f\right)\right)\Big/\sum_{i=1}^{N_2}\rho_{SM,i}\,g_i\left(\boldsymbol{M}_{Sx}\left(n_f\right)\right) \tag{3-7}$$

where $G\left(i \mid \boldsymbol{P}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SP}\right)$ and $G\left(i \mid \boldsymbol{M}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SM}\right)$ are *a posteriori* probabilities.

Maximization step:

(i). Estimate the mixture weights:

$$\rho_{SP,i}=1/N_F\sum_{n_f=1}^{N_F}G\left(i \mid \boldsymbol{P}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SP}\right) \tag{3-8}$$

$$\rho_{SM,i}=1/N_F\sum_{n_f=1}^{N_F}G\left(i \mid \boldsymbol{M}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SM}\right) \tag{3-9}$$

(ii). Estimate the mean vector:

$$\boldsymbol{\mu}_{SP,i}=\sum\nolimits_{n_f=1}^{N_F}G\left(i \mid \boldsymbol{P}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SP}\right)\boldsymbol{P}_{Sx}\left(n_f\right)\Big/\sum\nolimits_{n_f=1}^{N_F}G\left(i \mid \boldsymbol{P}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SP}\right) \tag{3-10}$$

$$\boldsymbol{\mu}_{SM,i}=\sum\nolimits_{n_f=1}^{N_F}G\left(i \mid \boldsymbol{M}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SM}\right)\boldsymbol{M}_{Sx}\left(n_f\right)\Big/\sum\nolimits_{n_f=1}^{N_F}G\left(i \mid \boldsymbol{M}_{Sx}\left(n_f\right),\boldsymbol{\lambda}_{SM}\right) \tag{3-11}$$

(iii). Estimate the variances:

$$\sigma^2_{SP,i}(\omega_b)=\left(\sum_{n_f=1}^{N_F} G\left(i\,|\,\boldsymbol{P}_{Sx}(n_f),\boldsymbol{\lambda}_{SP}\right)P_{Sx}{}^2(n_f,\omega_b)\Big/\sum_{n_f=1}^{N_F} G\left(i\,|\,\boldsymbol{P}_{Sx}(n_f),\boldsymbol{\lambda}_{SP}\right)\right)-\mu_{SP,i}{}^2(\omega_b) \tag{3-12}$$

$$\sigma^2_{SM,i}(\omega_b)=\left(\sum_{n_f=1}^{N_F} G\left(i\,|\,\boldsymbol{M}_{Sx}(n_f),\boldsymbol{\lambda}_{SM}\right)M_{Sx}{}^2(n_f,\omega_b)\Big/\sum_{n_f=1}^{N_F} G\left(i\,|\,\boldsymbol{M}_{Sx}(n_f),\boldsymbol{\lambda}_{SM}\right)\right)-\mu_{SM,i}{}^2(\omega_b) \tag{3-13}$$

An accelerated K-means algorithm proposed in [55] is again utilized to reduce the computational power requirement.

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

$$F_{RLM}(l)=\alpha_{LP}G\left(\boldsymbol{P}_{Lx}(n_f)\,|\,\boldsymbol{\lambda}_{LP}(l)\right)+\alpha_{LM}G\left(\boldsymbol{M}_{Lx}(n_f)\,|\,\boldsymbol{\lambda}_{LM}(l)\right) \tag{3-14}$$

$$F_{ROM}(o)=\alpha_{OP}G\left(\boldsymbol{P}_{Ox}(n_f)\,|\,\boldsymbol{\lambda}_{OP}(o)\right)+\alpha_{OM}G\left(\boldsymbol{M}_{Ox}(n_f)\,|\,\boldsymbol{\lambda}_{OM}(o)\right) \tag{3-15}$$

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:

$$\min\left\{\sum_{\boldsymbol{q}_{SP}}\alpha_{SP}\left\{\mathbf{C}_{SP}(\boldsymbol{q}_{SP})\mathbf{U}\mathbf{C}_{SP}(\boldsymbol{q}_{SP})^{T}\right\}+\sum_{\boldsymbol{q}_{SM}}\alpha_{SM}\left\{\mathbf{C}_{SM}(\boldsymbol{q}_{SM})\mathbf{U}\mathbf{C}_{SM}(\boldsymbol{q}_{SM})^{T}\right\}\right\} \tag{3-16}$$

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

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

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

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,

$$C(\boldsymbol{q}_{SP}\,|\,\boldsymbol{\lambda}_{SP}(l))=\frac{H(\boldsymbol{q}_{SP}\,|\,\boldsymbol{\lambda}_{SP}(l))}{\sqrt{\sum_{\boldsymbol{q}_{SP}}H^{2}(\boldsymbol{q}_{SP}\,|\,\boldsymbol{\lambda}_{SP}(l))}}\quad,$$

$$C(\boldsymbol{q}_{SM}\,|\,\boldsymbol{\lambda}_{SM}(l))=\frac{H(\boldsymbol{q}_{SM}\,|\,\boldsymbol{\lambda}_{SM}(l))}{\sqrt{\sum_{\boldsymbol{q}_{SM}}H^{2}(\boldsymbol{q}_{SM}\,|\,\boldsymbol{\lambda}_{SM}(l))}},$$

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

$$H(\boldsymbol{q}_{SM} \mid \boldsymbol{\lambda}_{SM}(l)) = G(\boldsymbol{q}_{SM} \mid \boldsymbol{\lambda}_{SM}(l)) - \frac{\sum_{\boldsymbol{q}_{SM}} G(\boldsymbol{q}_{SM} \mid \boldsymbol{\lambda}_{SM}(l))}{N(\boldsymbol{q}_{SM})}$$

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

The values of α_{SP} and α_{SM} can be obtained by solving (3-16) as:

$$\alpha_{SP} = \sqrt{\frac{\sum_{\boldsymbol{q}_{SM}} \mathbf{C}_{SM}(\boldsymbol{q}_{SM})\mathbf{U}\mathbf{C}_{SM}(\boldsymbol{q}_{SM})^T}{\sum_{\boldsymbol{q}_{SP}} \mathbf{C}_{SP}(\boldsymbol{q}_{SP})\mathbf{U}\mathbf{C}_{SP}(\boldsymbol{q}_{SP})^T}} \tag{3-17}$$

$$\alpha_{SM} = \sqrt{\frac{\sum_{\boldsymbol{q}_{SP}} \mathbf{C}_{SP}(\boldsymbol{q}_{SP})\mathbf{U}\mathbf{C}_{SP}(\boldsymbol{q}_{SP})^T}{\sum_{\boldsymbol{q}_{SM}} \mathbf{C}_{SM}(\boldsymbol{q}_{SM})\mathbf{U}\mathbf{C}_{SM}(\boldsymbol{q}_{SM})^T}} \tag{3-18}$$

3.4.2 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:

$$\begin{aligned}\hat{l} &= \arg\max_{1 \le l \le L} F_{RLM}(l) = \arg\max_{1 \le l \le L} \alpha_{LP} G(\boldsymbol{\lambda}_{LP}(l) \mid \mathbf{P}_{LY}) + \alpha_{LM} G(\boldsymbol{\lambda}_{LM}(l) \mid \mathbf{M}_{LY}) \\ &= \arg\max_{1 \le l \le L} \alpha_{LP} \frac{G(\mathbf{P}_{LY} \mid \boldsymbol{\lambda}_{LP}(l)) p(\boldsymbol{\lambda}_{LP}(l))}{p(\mathbf{P}_{LY})} + \alpha_{LM} \frac{G(\mathbf{M}_{LY} \mid \boldsymbol{\lambda}_{LM}(l)) p(\boldsymbol{\lambda}_{LM}(l))}{p(\mathbf{M}_{LY})}\end{aligned}$$

(3-19)

$$\hat{o} = \arg\max_{1\le o\le O} F_{ROM}(o) \ = \arg\max_{1\le o\le O} \alpha_{OP} \frac{G(\mathbf{P}_{OY} \mid \boldsymbol{\lambda}_{OP}(o))p(\boldsymbol{\lambda}_{OP}(o))}{p(\mathbf{P}_{OY})} + \alpha_{OM} \frac{G(\mathbf{M}_{OY} \mid \boldsymbol{\lambda}_{OM}(o))p(\boldsymbol{\lambda}_{OM}(o))}{p(\mathbf{M}_{OY})} \tag{3-20}$$

where $\mathbf{P}_{SY} = \{\boldsymbol{P}_{SY}(1),\cdots,\boldsymbol{P}_{SY}(N_V)\}$ and $\mathbf{M}_{SY} = \{\boldsymbol{M}_{SY}(1),\cdots,\boldsymbol{M}_{SY}(N_V)\}$ are the phase difference and magnitude ratio computed from the testing sequences denoted as $Y_{S1}(\omega)$ and $Y_{S2}(\omega)$, and N_V denotes the testing sequence length. The probabilities $p(\boldsymbol{\lambda}_{LP}(l))$ and $p(\boldsymbol{\lambda}_{LM}(l))$ can be selected as $1/L$ and $p(\boldsymbol{\lambda}_{OP}(o))$ and $p(\boldsymbol{\lambda}_{OM}(o))$ can 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 $p(\mathbf{P}_{SY})$ and $p(\mathbf{M}_{SY})$ are the same for all location models, the detection rule can be recast as:

$$\hat{l} = \arg\max_{1\le l\le L} \alpha_{LP} \prod_{n_v=1}^{N_V} G(\boldsymbol{P}_{LY}(n_v) \mid \boldsymbol{\lambda}_{LP}(l)) + \alpha_{LM} \prod_{n_v=1}^{N_V} G(\boldsymbol{M}_{LY}(n_v) \mid \boldsymbol{\lambda}_{LM}(l)) \tag{3-21}$$

$$\hat{o} = \arg\max_{1\le o\le O} \alpha_{OP} \prod_{n_v=1}^{N_V} G(\boldsymbol{P}_{OY}(n_f) \mid \boldsymbol{\lambda}_{OP}(o)) + \alpha_{OM} \prod_{n_v=1}^{N_V} G(\boldsymbol{M}_{OY}(n_f) \mid \boldsymbol{\lambda}_{OM}(o)) \tag{3-22}$$

3.5 Summary

A robot's location and orientation detection method based on sound field features utilizing two microphones is proposed in this chapter. The proposed method treats phase difference and magnitude ratio distributions between the microphones as sound field features. Based on this idea, RLMs and ROMs are introduced for robot's location and orientation detection. The system architecture presented contains a RLODA that can provide adaptation to environmental noises. Moreover, with the pre-recorded database, the non-ideal issues of non-line-of-sight condition and microphone mismatch problem can be solved. The related experimental results are shown in Chapter 4.

Chapter 4

Experimental Results

4.1 Experimental Results of the Proposed GMBRDM

4.1.1 The Experimental Environment

The experiment is performed in a laboratory filled with common furniture and equipment. Fig. 4-1 shows the layout of the environment. The laboratory area is $10.5\times7.1\ \mathrm{m}^2$ and room height is 3 m. The recording equipment comprises two B&K 4935 array microphones, a B&K 2694 conditioning amplifier, and an Azova DAQP-16 analog-to-digital converter.

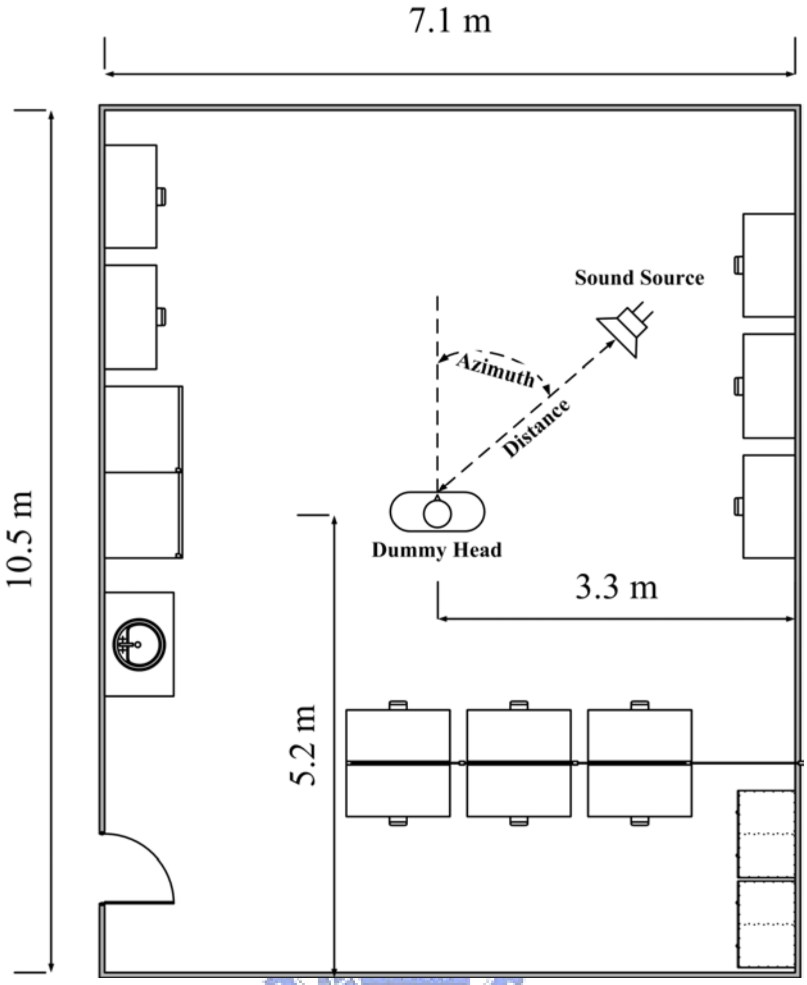


Figure 4-1 The layout of the experimental environment.

The microphones are mounted in the ears of a dummy head, as depicted in Fig. 4-2. The distance between the dummy head's ears is 0.16 m. Fig. 4-1 illustrates the location of the dummy head. The ears of the dummy head are placed 1 m above the floor.

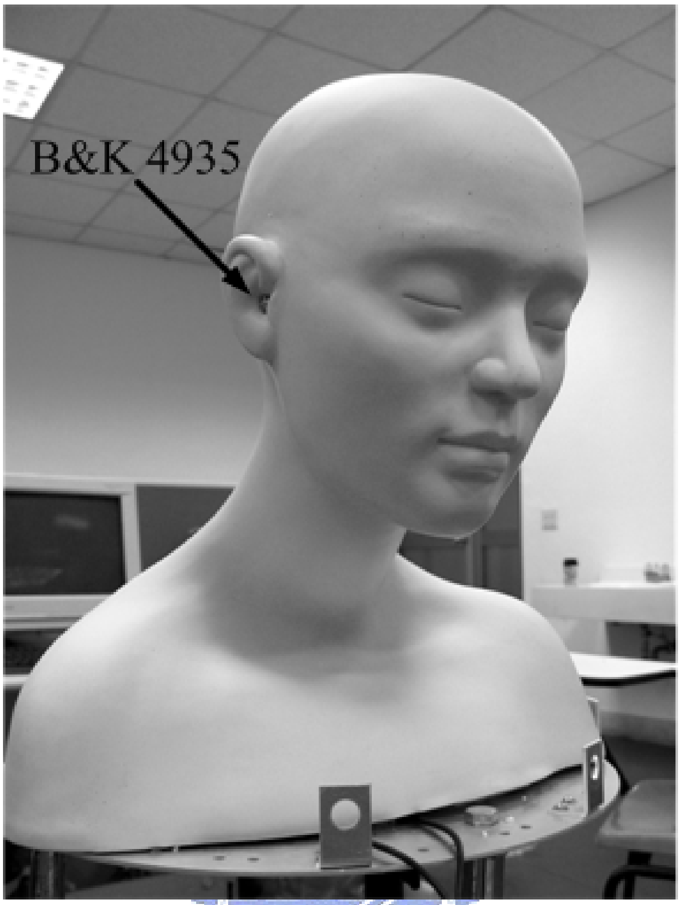


Figure 4-2 The dummy head adopted in the experiment.

The sound source is a recording of a female reading a book in Mandarin (Here, we assume the distribution of a_1 of the speech signal is stationary during training and testing procedure.). The sound source is played by a loudspeaker. Received signals are sampled at 8000 Hz, and the STFT window is 512 samples. For each experiment, the sound source is played at each tested location to obtain the training sequence to establish the GMBRDM. Training sequence length, N_F, is set to 400 and testing sequence length, N_V, is set to 100, with a shift of 80 samples between each frame. Hence, 4-second data are utilized for training, and 1-second data are utilized for testing. Six significant frequencies of the sound source are selected in this experiment; therefore, each Gaussian model has six dimensions, $B = 6$. For each location, testing is performed 100 times to acquire the correct rate.

4.1.2 The Experimental Results

The first experiment tests the ability of azimuth localization. In this experiment, distances between the sound source and ears are fixed at 1 m, 1.2 m, 1.4 m, 1.6 m, 1.8 m, and 2 m. For each distance, the azimuth of sound source moves from -60°, -30°, 0°, 30°, to 60° to test the average correct rate of azimuth localization. The elevation of sound source is set the same as that of the ears (1 m). Different mixture numbers are utilized. Table 4-1 shows the average correct rate of azimuth localization at each distance.

Table 4-1 Average correct rates of azimuth localization at each distance

Mixture	Distance (m)					
Number	1.0	1.2	1.4	1.6	1.8	2.0
1	97 %	83 %	40 %	67 %	70 %	67 %
5	98 %	84 %	59 %	81 %	85 %	72 %
10	99 %	89 %	81 %	87 %	88 %	73 %
15	99 %	91 %	83 %	87 %	89 %	83 %
20	99 %	88 %	83 %	86 %	89 %	85 %
25	99 %	91 %	88 %	87 %	89 %	91 %

As shown in Table 4-1, when the distance between the sound source and ears is 1 m, meaning that the sound source close to the dummy head, the performance of mixture number of 1 is roughly the same as those of high mixture numbers. When the sound source is close to the dummy head, the influence of direct path propagation is much more significant than that of reverberations. Consequently, the BRDPs are influenced less by the reflections and can be modeled using a single Gaussian distribution model. However, as distance between the sound source and ears increases, the influence of reflection is becoming significant and leads to complex BRDPs. The benefit of adopting multiple mixtures is apparent at a long distance, such as 2 m, where the correct rate increases with the mixture number.

The second experiment tests the capability of the proposed GMBRDM for distance localization. In this experiment, the azimuth is fixed at -60°, -30°, 0°, 30°, and 60°. At each azimuth, the distance between the sound source and ears changes from 1 m, 1.2 m, 1.4 m, 1.6 m, 1.8 m, to 2 m to acquire average correct rates. The sound source height is adjusted to 1 m. Table 4-2 shows the average correct rates for distance localization at each azimuth.

Table 4-2 Average correct rates of distance localization at each azimuth

Mixture	Azimuth				
Number	-60°	-30°	0°	30°	60°
1	49 %	31 %	48 %	43 %	61 %
5	40 %	47 %	65 %	55 %	64 %
10	76 %	68 %	73 %	58 %	69 %
15	80 %	76 %	73 %	67 %	72 %
20	79 %	73 %	73 %	70 %	74 %
25	86 %	82 %	73 %	73 %	78 %

Because the relationship between the sound source and ears meets the criterion of far-field, the IPDs of direct path at the same azimuth and different distances are approximately identical theoretically. The ILDs of direct paths generate only relatively a slight difference between distant locations. Thus, modeling these BRDPs using a single Gaussian component can lose important details caused by reflections and result in poor localization results. As listed in Table 4-2, the average correct rates when only one mixture is employed are clearly lower than those with a high mixture number. This experimental finding is because the proposed GMBRDM can represent the details of the BRDPs for superior modeling results.

The third experiment tests the elevation localization performance of the proposed GMBRDM. In this experiment, distance between the sound source and ears is 2 m and the azimuth is fixed at -60°, -30°, 0°, 30°, and 60°. At each azimuth, the elevation of the sound source changes from 1 m, 1.25 m, to 1.5 m to acquire average correct rates. Table 4-3 lists experimental results. Experimental data show that GMBRDM with a high mixture number can properly model the BRDPs at different elevations.

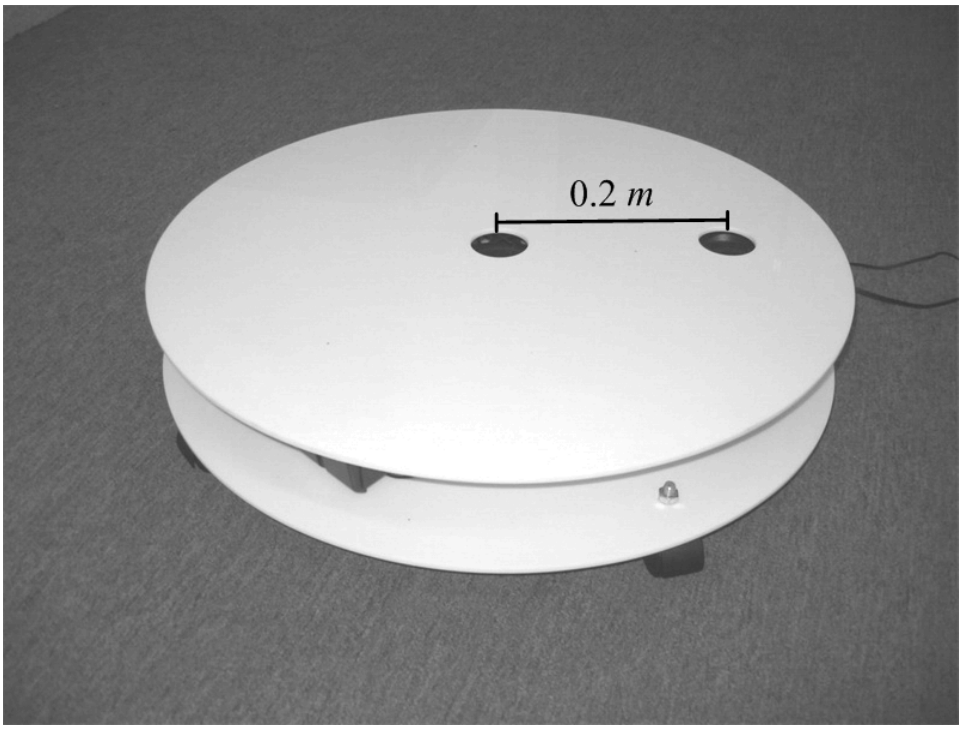
Table 4-3 Average correct rates of elevation localization at each azimuth

Mixture	Azimuth				
Number	-60°	-30°	0°	30°	60°
1	59 %	55 %	33 %	46 %	59 %
5	83 %	90 %	60 %	80 %	63 %
10	82 %	93 %	55 %	83 %	74 %
15	88 %	94 %	55 %	88 %	74 %
20	89 %	93 %	67 %	86 %	81 %
25	92 %	98 %	84 %	93 %	88 %

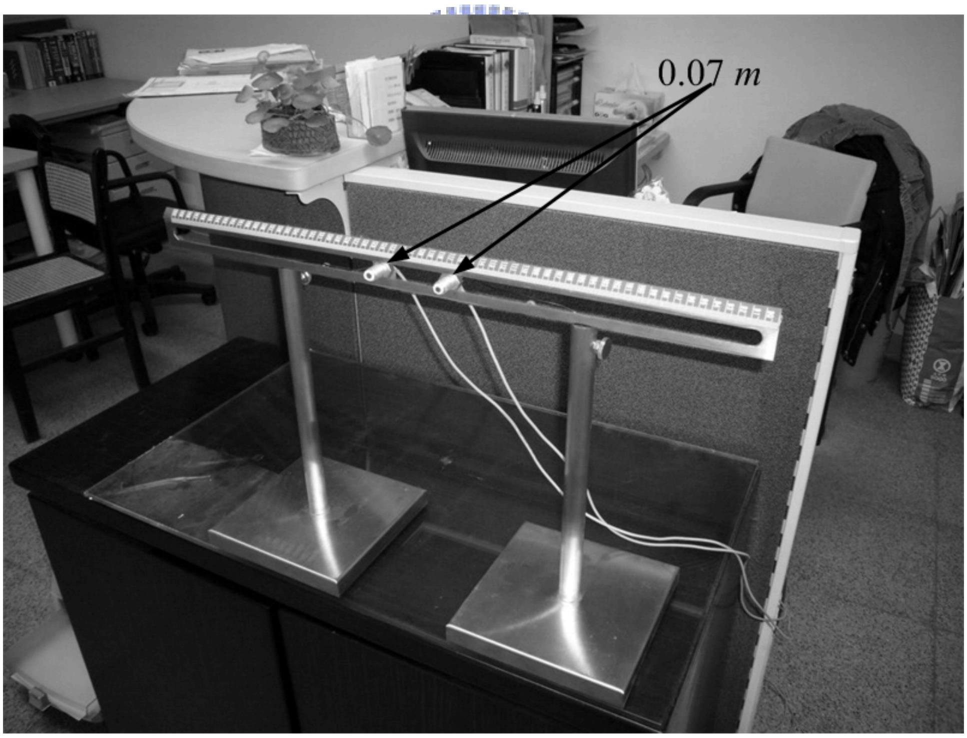
4.2 Experimental Results of the Proposed Robot’s Localization and Orientation Detection Method

4.2.1 The Experimental Environment

Figure 4-3 shows the experimental platform and the proposed RLODA. In Fig. 4-3 (a), the distance between two loudspeakers is 0.2 m. The distance between the two microphones of the RLODA is chosen as 0.07 m, as shown in Fig. 4-3 (b).



(a)



(b)

Figure 4-3 The experimental platform and the proposed RLODA. (a) The experimental platform. (b) The proposed RLODA.

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 8000 Hz, and the A/D resolution is 16 bits. 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°.

Figure 4-4 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. 4-4 with the dotted lines from 1 to 8 in sequence.

The sound source utilized in this experiment mimic the sound of dog barking. The spectrogram of the sound source is illustrated in Fig. 4-5. 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.

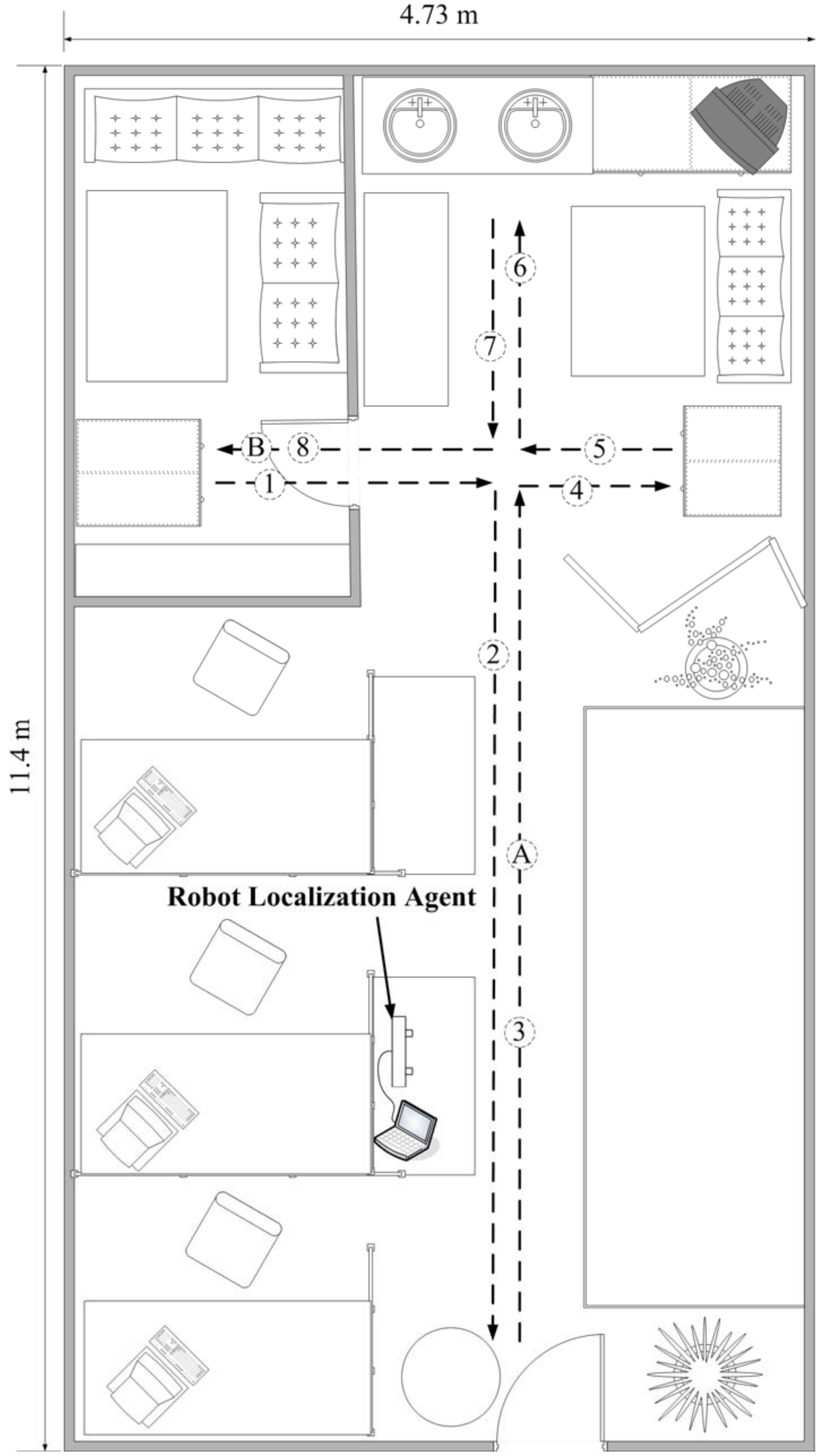


Figure 4-4 Experimental environment.

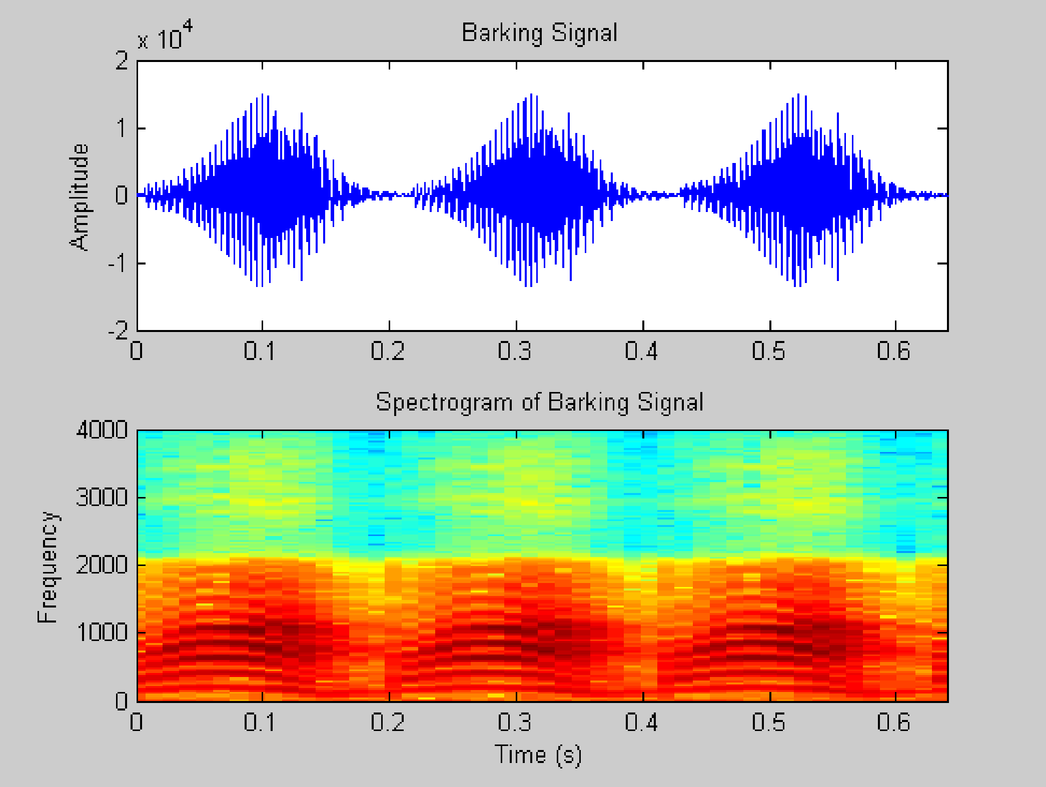


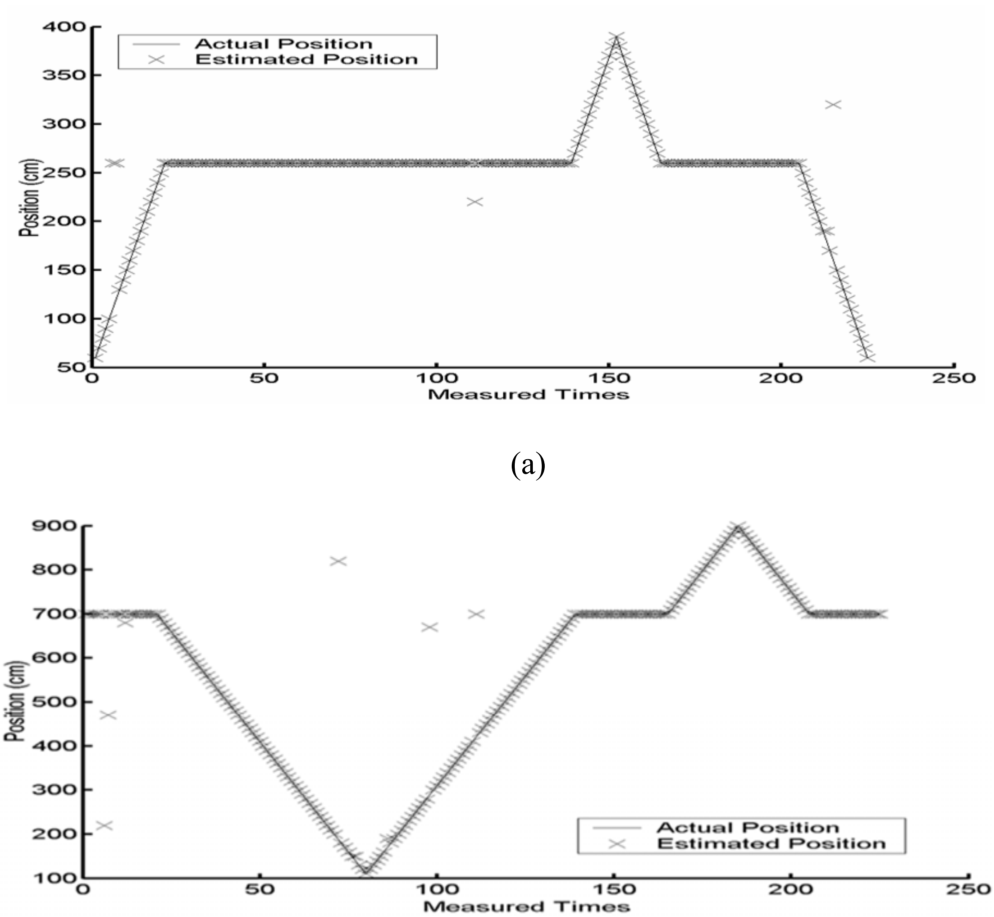
Figure 4-5 The waveform and spectrogram of barking signal.

4.2.2 The Experimental Results

Table 4-4 lists the average SNRs of all trajectories and the average SNRs of each trajectory pair. Figure 4-6 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. 4-6, the location detection results are mostly very close to the actual location for most of the time.

Table 4-4 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



(a)

(b)

Figure 4-6 Location detection results alone X and Y axes. (a) Location detection results alone X-axis. (b) Location detection results alone Y-axis.

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. 4-6, 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 4-4. 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.

Several experiments are conducted to access the accuracy of the proposed method in terms of location and orientation detection error. Table 4-5 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 4-5, 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 $0 \le D < 1\,cm$.

Table 4-5 Average correct rates of location detection results (%)

Average SNR (dB)	M = 1			M = 11			M = 15		
	$0 \le D < 1$ (cm)	$1 \le D < 3$ (cm)	$3 \le D < 5$ (cm)	$0 \le D < 1$ (cm)	$1 \le D < 3$ (cm)	$3 \le D < 5$ (cm)	$0 \le D < 1$ (cm)	$1 \le D < 3$ (cm)	$3 \le D < 5$ (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 4-6 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 4-6, 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 $0^\circ \le A < 4^\circ$, the average correct rates exceed 99% in both the SNR cases.

Table 4-6 Average correct rates of orientation detection results (%)

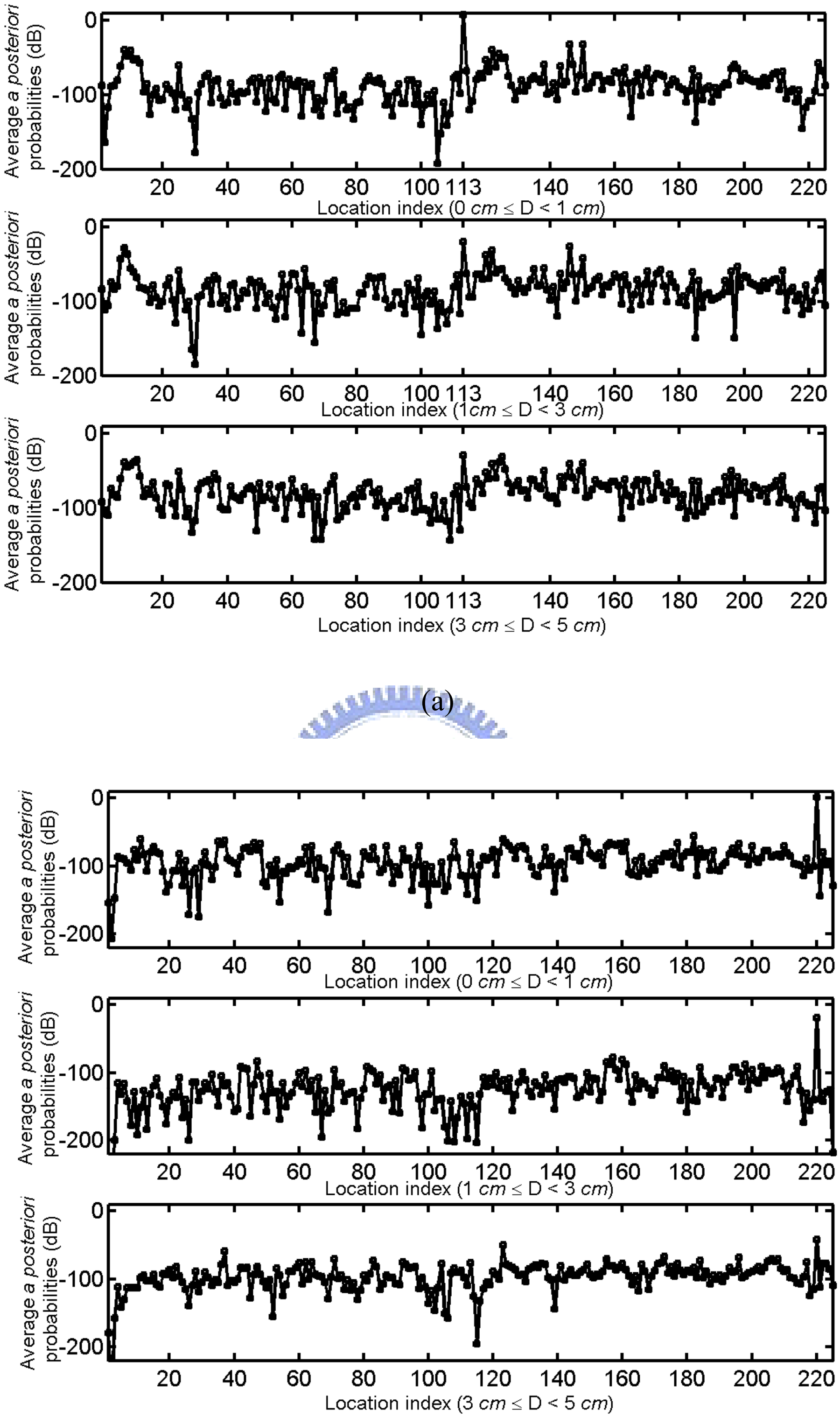
Average SNR (dB)	M = 1				M = 11			
	$0^\circ \le A < 4^\circ$	$4^\circ \le A < 8^\circ$	$8^\circ \le A < 12^\circ$	$12^\circ \le A < 15^\circ$	$0^\circ \le A < 4^\circ$	$4^\circ \le A < 8^\circ$	$8^\circ \le A < 12^\circ$	$12^\circ \le A < 15^\circ$
19.16	58.43	48.49	45.78	44.28	99.70	88.55	84.04	81.33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

Figure 4-7 and 4-8 show the average of *a posteriori* probabilities measured at the locations “A” and “B”, where location “A” is in a line-of-sight case and location “B” is in a non-line-of-sight case, as illustrated in Fig. 4-4. Notably, the *a posteriori* location probability is defined as:

$$\alpha_{LP}\prod_{n_v=1}^{N_V} G\left(\boldsymbol{P}_{LY}(n_f)\,|\,\boldsymbol{\lambda}_{LP}(l)\right)+\alpha_{LM}\prod_{n_v=1}^{N_V} G\left(\boldsymbol{M}_{LY}(n_f)\,|\,\boldsymbol{\lambda}_{LM}(l)\right) \tag{4-1}$$

and the *a posteriori* orientation probability is defined as:

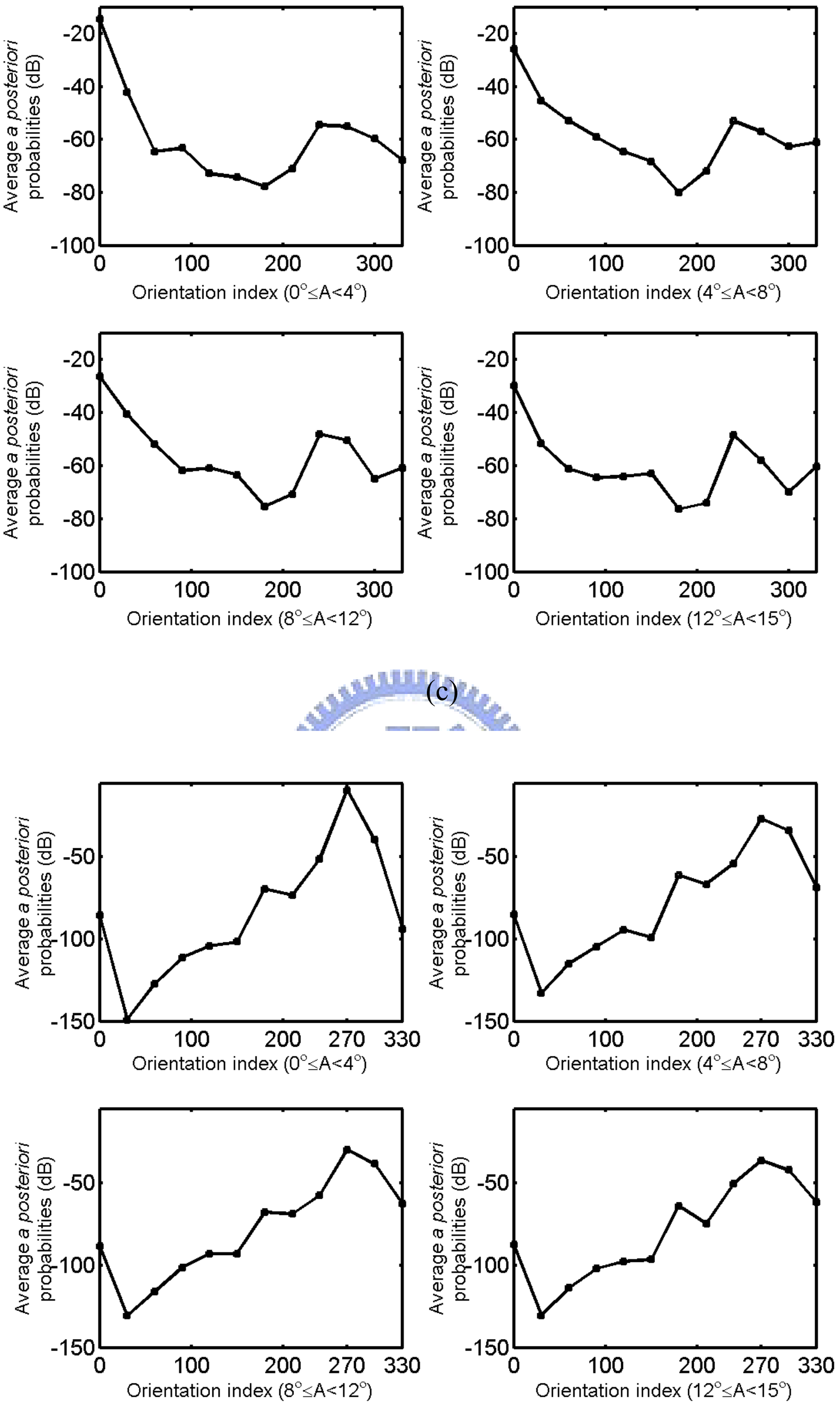
$$\alpha_{OP}\prod_{n_v=1}^{N_V} G\left(\boldsymbol{P}_{OY}(n_f)\,|\,\boldsymbol{\lambda}_{OP}(l)\right)+\alpha_{OM}\prod_{n_v=1}^{N_V} G\left(\boldsymbol{M}_{OY}(n_f)\,|\,\boldsymbol{\lambda}_{OM}(l)\right) \tag{4-2}$$



(a)

(b)

Figure 4-7 The average of the measured *a posteriori* location probabilities. (a) The average *a posteriori* location probabilities at location "A". (b) The average *a posteriori* location probabilities at location "B".



(c)

(d)

Figure 4-8 The average of the measured *a posteriori* orientation probabilities. (a) The average *a posteriori* orientation probabilities at location "A". (b) The average *a posteriori* orientation probabilities at location "B".

The average SNRs belong to the lowest SNR conditions in Table 4-5 and Table 4-6 individually. The mixture number in Fig. 4-7 is 15, and the mixture number in Fig. 4-8 is 11. Location "A" denotes the 113th location and location "B" represents the 220th location. In the case of $0\,cm \le D < 1\,cm$, the averages of (4-1) (averages of *a posteriori* location probabilities) measured with the correct locations indices ($l = 113$ and $l = 220$) are much higher than those of other location indices, as shown in Fig. 4-7 (a) and (b). However, since the sound field feature varies with the robot'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. 4-7 (a) and (b), the difference between the averages of (4-1) 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 4-5 degrade with the increase of the distances between the actual and the pre-recorded locations. Although the averages of (4-1) 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 4-5 remain above 80% when $3\,cm \le D < 5\,cm$.

The same phenomenon appears in the experiment of orientation detection. Figure 4-8 (a) and (b) depict the average of (4-2) (averages of *a posteriori* orientation probabilities) with the correct orientations of 0° for Fig. 4-8 (a) and 270° for Fig. 4-8 (b). The average of (4-2) 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 4-6. These experimental results in this section show that utilizing GMMs to model the sound field features is a feasible method for robot’s location and orientation detection.

Chapter 5

Conclusions and Potential Research Topics

5.1 Conclusions

This dissertation has investigateed the relationship between nonstationary sound sources and the BRDPs when STFT is utilized. First, the level fluctuation of the nonstationary sound source is modeled as an exponent of polynomial based on the concept of moving pole model. This model explains the content dependency of the BRDPs. Moreover, the sufficient condition for utilizing BRDPs to detect the location of nonstationary sound source is identified. The phenomena of multiple peaks in the distribution patterns are analyzed. The related derivation shows that using simple distribution, such as a single Gaussian distribution, is not suitable for modeling these distribution patterns. Therefore, a GMBRDM is proposed to model the BRDPs for nonstationary sound source localization. Experimental results display that the proposed GMBRDM can discriminate between the azimuth, elevation, and distance of the sound sources. Notably, the correct rates in experimental results do not monotonically increase with the number of Gaussian mixtures. This experimental finding is because the proposed GMBRDM can be influenced by the initial condition selected and the complexity of BRDPs varies with sound source locations.

Moreover, a novel robot's location and orientation detection method based on sound field features utilizing two microphones 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. 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.

5.2 Potential Research Topics

5.2.1 The Prediction, Interpolation, or Extrapolation of BRDPs

The relation between the BRDPs and the location of sound source can only be obtained by empirical data at present state. However, the demand of empirical data can restrict the application of the proposed localization methods. Consequently, the studies of room acoustic channel modeling can be combined with this research to solve this problem and provide a wider application scenario.

Sometimes, the BRDPs of some points are known in advance. However, researchers might be interested in the BRDPs at the locations between or near these points. Therefore, the interpolation or extrapolation of BRDPs is also an important research topic.

5.2.2 The Influence of Environmental Change to the BRDPs

The proposed localization methods assume that the BRDPs or sound field features remain unchanged during training and testing procedures. Nevertheless, in real environment, the configuration of the room can alter with time. The influence of the environmental change has been roughly discussed in [49]. However, the work in [49] only considered the basic geometry change; the detail of the influence needs further explore. Moreover, the variation of temperature would influence the sound speed; hence results in difference of propagation model. Therefore, besides the change of the configuration of the room, variation of temperature would also alter the measured BRDPs. The relation between the temperature and the room acoustic transfer function has been discussed in [72]. The related research can be adopted to compensate the effect of temperature variation.

5.2.3 Robot's Location and Orientation Detection Using Hidden Markov Model

The proposed RLM and ROM utilize only the currently measured IPDs and ILDs for location and orientation detection. However, for some application, if the relative displacement of the sound source or the robot is known, the previously measured IPDs and ILDs also provide important information. The locations and orientations can be treated as "states" in the hidden Markov model (HMM) [53]. Therefore, instead detecting the most possible location of the robot by the proposed RLMs and ROMs, using HMM can detect the most possible trajectory of the robot. Furthermore, the state transition matrix in the HMM can be utilized to combine with other sensors or localization methods.

References

[1] M. Brandstein and H. Silverman, "A Practical Methodology for Speech Source Localiration with Microphone arrays," *Compurer, Speech, and Language*, vol. 11, no.2, pp. 91-126, Apr. 1997.

[2] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *Journal of Acoustical Society of America*, vol. 85, pp. 858-867, Feb. 1989.

[3] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *Journal of Acoustical Society of America*, vol. 85, pp. 868-878, Feb. 1989.

[4] S. Carlile, *Virtual auditory space: Generation and application*, New York: Chapman and Hall, 1996.

[5] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *Journal of Acoustical Society of America*, vol. 97, no. 6, pp. 3907-3909, June 1995.

[6] H. S. Colburn and A. Kulkarni, "Models of sound localization," *in Sound Source Localization, R. Fay and T. Popper, Eds., Springer Handbook of Auditory Research,* Springer-Verlag, 2005.

[7] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," *Annu. Rev. Psychol.*, vol. 42, pp. 135-159, Jan. 1991.

[8] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz, "Auditory localization of nearby sources. II. Localization of a broadband source," *Journal of Acoustical Society of America*, vol. 106, no. 4, pp. 1956-1968, Oct. 1999.

[9] C. Trahiotis, L. R. Bernstein, R. M. Stern, and T. N. Buell, "Interaural correlation as the basis of a working model of binaural processing: an introduction," *in Sound Source Localization, R. Fay and T. Popper, Eds., Springer Handbook of Auditory Research*, Springer-Verlag, 2005.

[10] C. H. Knapp, and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustic, Speech, Signal Processing*, vol. 24, pp. 320-327, Aug. 1976.

[11] G. C. Carter, A. H. Nuttall, and P. G. Cable, "The smoothed coherence transform," *IEEE Signal Processing Letters*, vol. 61, pp. 1497-1498, Oct. 1973.

[12] R. S. Woodworth, *Experimental psychology*, New York Halt, 1938.

[13] P. M. Hofman and A. J. von Opstal, "Spectro-temporal factors in two-dimensional

human sound localization," *Journal of Acoustical Society of America*, vol. 103, no. 5, pp. 2634-2648, May 1998.

[14] J. P. Blauert, *Spatial Hearing*, MIT Prees, Cambridge, MA, 1983.

[15] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *Journal of Acoustical Society of America*, vol. 109, no. 3, pp. 1110-1122, Mar. 2001.

[16] P. Zakarouskas and M. S. Cynader, "A computational theory of spectral cue localization," *Journal of Acoustical Society of America*, vol. 94, no. 3, pp. 1323-1331, Sept. 1993.

[17] J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics," *Journal of Acoustical Society of America*, vol. 92, no. 5, pp. 2607-2624, Nov. 1992.

[18] B. G. Shinn-Cunningham, "Distance cues for virtual auditory space," *Proceedings of IEEE Conference on Multimedia*, pp. 227-230, Dec. 2000

[19] J. G. Ryan, R. A. Goubran, "Near-field beamforming for microphone arrays," *IEEE Internationa Conference on Acoustics, Speech, and Signal Processing,* vol. 1, pp. 21-24, Apr. 1997.

[20] M. Wax and T. Kailath, "Optimal localization of multiple sources by passive arrays," *IEEE Transactions on Acoustic, Speech, Signal Processing*, vol. ASSP-31, pp. 1210-1217, Oct. 1983.

[21] H. F. Silverman and S. E. Kirtman, "A two-stage algorithm for determining talker location from linear microphone-array data," *Computer, Speech, and Language*, vol. 6, pp. 129-152, Apr. 1992.

[22] D. B. Ward, and R. C. Williamson, "Particle filter beamforming for acoustic source localization in a reverberant environment," *IEEE International Conference on Acoustics, Speech, and Signal Processing,* vol. 2, pp. 1777-1780, May 2002.

[23] R. V. Balan and J. Rosca, "Apparatus and method for estimating the direction of Arrival of a source signal using a microphone array," *European Patent, No US2004013275,* 2004.

[24] M. Wax, T. J. Shan, and T. Kailath. " Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Transactions on Acoustic Speech, Signal Processing*, vol ASSP-32, pp. 817–827, Aug. 1984.

[25] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation,* vol. 34, pp. 276-280. Mar. 1986.

[26] H. Wang and M. Kaveh, "Coherent signal subspace processing for detection and estimation of angle of arrival of multiple wideband sources," *IEEE Transactions on Acoustic Speech, Signal Processing*, vol ASSP-33, pp. 823–831, Aug. 1985.

[27] G. Bienvenu, "Eigensystem properties of the sampled space correlation matrix,"

IEEE International Conference on Acoustic, Speech, and Signal Processing, pp. 332–335, Apr. 1983.

[28] K. M. Buckley and L. J. Griffiths, "Eigenstructure based broadband source location estimation," *IEEE International Conference on Acoustic, Speech, and Signal Processing*, pp. 1869–1872, Apr. 1986.

[29] M. A. Doron, A. J. Weiss, and H. Messer, "Maximum likelihood direction finding of wideband sources," *IEEE Transactions on Signal Processing,* vol. 41, pp. 411–414, Jan. 1993.

[30] M. Agarwal and S. Prasad, "DOA estimation of wideband sources using a harmonic source model and uniform linear array," *IEEE Transactions on Signal Processing*, vol. 47, pp. 619-629, Mar. 1999.

[31] H. Messer, "The potential performance gain in using spectral information in passive detection/localization of wideband sources," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2964-2974, Dec. 1995.

[32] M. Agrawal and S. Prasad, "Broadband DOA estimation using spatial-only modeling of array data," *IEEE Transactions on Signal Processing*, vol. 48, pp. 663-670, Mar. 2000.

[33] J. H. Lee, Y. M. Chen, and C. C. Yeh, "A covariance approximation method for near-field direction finding using a uniform linear array," *IEEE Transactions on Signal Processing*, vol. 43, pp. 1293-1298, May 1995.

[34] K. Buckley and L. Griffiths, "Broad-band signal-subspace spatial-spectrum (BASS-ALE) estimation," *IEEE Transactions on Acoustic, Speech, Signal Processing*, vol. 36, pp. 953-964, July 1988.

[35] N. Strobel and R. Rabenstein, "Classification of time delay estimates for robust speaker localization," *IEEE International Conference on Acoustics, Speech, and Signal Processing,* vol. 6, pp. 15-19, Mar. 1999.

[36] J. S. Hu, T. M. Su, C. C. Cheng, W. H. Liu, and T. I. Wu, "A self-calibrated speaker tracking system using both audio and video data," *IEEE Conference on Control Applications*, vol.2, pp. 731-735, Sept. 2002.

[37] J. S. Hu, C. C. Cheng, W. H. Liu, and T. M. Su, "A speaker tracking system with distance estimation using microphone array," *IEEE/ASME International Conference on Advanced Manufacturing Technologies and Education,* Aug. 2002.

[38] S. Mavandadi, P. Aarabi, "Multichannel nonlinear phase analysis for time-frequency data fusion," *Proceedings of the SPIE, Architectures, Algorithms, and Applications VII (AeroSense 2003),* vol. 5099, pp. 222-231, Apr. 2003.

[39] P. Aarabi and S. Mavandadi, "Robust sound localization using conditional time–frequency histograms," *Information Fusion*, vol. 4, pp. 111-122, June 2003.

[40] D. D. Rife and J. Vanderkooy, "Transfer-function measurement using maximum-length sequences," *Journal of Acoustical Society of America*, vol. 37, no. 6, pp. 419-444, June 1989.

[41] W. M. Hartmann, "Localization of sound in rooms," *Journal of Acoustical Society of America*, vol. 74, no. 5, pp. 1380-1391, Nov. 1983.

[42] H. Kuttruf, *Room acoustics*. London: Elsevier, 1991, chapter 3, pp. 56.

[43] T. Gustafsson, B. D. Rao, and M. Trivedi, "Source localization in reverberant environments: modeling and statistical analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 791-803, Aug. 2003.

[44] B. G. Shinn-Cunningham, N. Kopcp and T. J. Martin, "Localizing nearby sound sources in a classroom: binaural room impulse response," *Journal of Acoustical Society of America*, vol. 117, no. 5, pp. 3100-3115, May 2005.

[45] B. G. Shinn-Cunningham, "Localizing sound in rooms," in *Proceedings of the ACM SIGGRAPH and EUROGRAPHICS Campfire: Rendering for Virtual Environments*, pp. 17-22, May 2001.

[46] J. Huang, N. Ohnishi, and N. Sugie, "Sound localization in reverberant environment based on the model of the precedence effect," *IEEE Transactions on Instrumentation and Measurement*, vol. 46, no. 4, pp. 842-846, Aug. 1997.

[47] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of Acoustic Society America* vol. 65, Issue 4, pp. 943-950, Apr. 1979.

[48] J. Nix and V. Hohmann, "Sound source localization in real sound fields based on empirical statistics of interaural parameters," *Journal of Acoustical Society of America*, vol. 119, no. 1, pp. 463-479, Jan. 2006.

[49] P, Smaragdis and P. Boufounos, "Position and trajectory learning for microphone arrays," *IEEE Transactions on Audio, Speech and Language Processing,* vol. 15, no. 1, pp. 358-368, Jan. 2007.

[50] Y. H. Tsao, "Tests for nonstationarity," *Journal of Acoustic Society America*, vol. 75, Issue 2, pp. 486-498, Feb. 1984.

[51] D. H. Friedman, "Estimation of formant parameters by sum-of-poles modeling," in *Proceedings of ICASSP*, pp. 351-354, Apr. 1981.

[52] F. Casacuberta and E. Vidal, "A nonstationary model for the analysis of transient speech signals" *IEEE Transactions Acoustic, Speech, and Signal Processing*, vol. ASSP-35, no. 2, pp. 226-228, Feb. 1987.

[53] G. Xuan, W. Zhang, and P. Chai, "EM algorithms of Gaussian mixture model and hidden Markov model," *IEEE International Conference on Image Processing*, pp. 145-148, Oct. 2001.

[54] J. B. MacQueen, "Some methods for classification and analysis of multivariate

observations", *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281-297, 1967.

[55] C. Elkan, "Using the triangle inequality to accelerate k-means," *Proceedings of the Twentieth International Conference on Machine Learning,* pp. 147-153, 2003.

[56] J. S. Hu, W. H. Liu, and C. C. Cheng, "Indoor sound field feature matching for robot's location and orientation detection," submitted to *Pattern Recognition Letters*.

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

[58] A. Georgiev, and P. K. Allen, "Localization methods for a mobile robot in urban environments," *IEEE Transactions on Robotics*, vol. 20, pp. 851-864, Oct. 2004.

[59] C. D. McGillem and T. S. Rappaport, "Infra-red location system for navigation of autonomous vehicles," *IEEE International Conference on Robotics and Automation*, pp. 1236-1238, Apr. 1988.

[60] 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 Transactions on Robotics and Automation*, vol. 14, no. 6, pp. 969-978, Dec. 1998.

[61] J. M. Lee, K. Son, M. C. Lee, J. W. Choi, S. H. Han, and M. H. Lee, "Localization of a mobile robot using the image of a moving robot," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 3, pp. 612-619, June 2003.

[62] R. Gutierrez-Osuna, J. A. Janet, and R. C. Luo, "Modeling of ultrasonic range sensors for localization of autonomous mobile robots," *IEEE Transactions on Industrial Electronics*, vol. 45, no. 4, pp. 654-662, Aug. 1998.

[63] U. Larsson, J. Frosberg, and A. Wernersson, "Mobile robot localization: integrating measurements from a time-of-flight laser," *IEEE Transactions on Industrial Electronics*, vol. 43, no. 3, pp. 422-431, June 1996.

[64] 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 Transactions on Robotics and Automation*, vol. 20, no. 3, pp. 555-559, June 2004.

[65] Q. H. Wang, T. Ivanov, and P. Aarabi, "Acoustic robot navigation using distributed microphone arrays," *Information Fusion*, Information Fusion 5, vol. 5, pp. 131-140, June 2004.

[66] Y. Tamai, S. Kagami, H. Mizoguchi, Y. Amemiya, K. Nagashima and T. Takano, "Real-time 2 dimensional sound source localization by 128-channel huge microphone array," *IEEE International workshop on Robot and Human Interactive Communication*, pp. 65-70, Sept. 2004.

[67] M. S. Brandstein and H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," *IEEE International Conference on*

Acoustics, Speech, and Signal Processing, vol. 1, pp. 375-378, Apr. 1997.

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

[69] Y. Tamai, S. Kagami, H. Mizoguchi, Y. Amemiya, K. Nagashima, and T. Takano, "Sound spot generation by 128-channel surround speaker array," *IEEE International workshop on Sensor array and multichannel signal processing*, pp. 542-546, July 2004.

[70] M. Yamada, N. Itsuki, and Y. Kinouchi, "Adaptive directivity control of speaker array," *Control, Automation, Robotics and Vision Conference*, pp. 1143-1148, Dec. 2004.

[71] S. P. Parker, *Acoustic Source Book*, McGraw-Hill, 1988.

[72] M. Omura, M. Yada, H. Saruwatari, S. Kajata, K. Takeda, and F. Itakura, "Compensation of room acoustic transfer function affected by change of room temperature," *IEEE International Conference on Acoustic, Speech, and Signal Processing*, pp. 941-944, Mar. 1999.