공학박사학위논문

Acoustic Sensor Localization Techniques Using Artificial Sound Sources in Reverberant Environments

잔향 환경에서의 인공 음향 신호를

이용한 음향 센서 위치 추정 기술

2017년 8월

서울대학교 대학원

전기·컴퓨터공학부

최 석 재

Abstract

Widespread use of smart devices has brought a growth of user-customized services. In particular, localization techniques have been gaining attention due to increase of location-based services (LBS). Most of LBS services such as navigation systems, traffic alerts or augmented reality (AR) services depend on the GPS for its accuracy and speed, however, its operation is limited to the outdoor environments. The demand of indoor LBS is rapidly growing due to the growth of automated home and IoT technology. There have been studies via WiFi, Bluetooth or RFID, but their performance has been unsatisfactory for their limitation such as the requirement of additional equipment or guarantee of the line of sight.

Among various sensors used for indoor localization, we focus on the acoustic sensors, i.e. microphones. There are several advantages in using the acoustic signals for indoor localization. There is no need for additional apparatus since loudspeakers are pre-installed in most of the buildings for the purpose of announcement or playing background music and mobile devices such as cellphones or tablets are equipped with microphones and loudspeakers. Even the prevailing popularity of IoT services helps accessibility of acoustical sensors and loudspeakers. In addition, acoustic signals have advantages of being able to detect signals through obstacles unlike cameras of RFID.

In this thesis, we propose a position estimation system using acoustic signals to

maximize these advantages. We aim to estimate the position of the target user with an acoustic sensor based on the recording of signals from the fixed loudspeakers installed around the room. We target to estimate the position of the acoustic sensor with high accuracy and low-complexity in a large space with high reverberation. Particularly, we try not to affect human hearing by using inaudible frequency bands. In order to estimate the position, it is important to estimate the direct path signal rather than the signal due to reverberation or reflection. To do this, we present various localization techniques as following.

First, we propose the source data structure to operate in the large reverberant environments. In the large space, the consideration of the near-far effect is required which refers to a situation when the desired signal is far away, it is difficult to receive the desired signal due to the interference of closer unwanted signals. In wireless communications, it can be dealt with by interaction of transmitter and receiver by feedback of channel information. However, it is difficult in the acoustic system since there is no feedback between the transmitter and receiver. We borrowed the structure called OFDMA-CDM and modified it to deal with the near-far effect. In the reverberant environment, the amplitude of reverberation is often larger than the direct path signal. We proposed the technique to estimate the direct path signal.

Second, we propose a method for accurate location estimation in the highly reverberant environments. In the high reverberation condition, more spurious reflections occur, which makes it difficult to estimate the time delay of the direct path signal. If the time delay estimation is wrong, it is likely that the position estimate does not converge by an estimation method. In the proposed method, position candidates are obtained from most of the received signals including signals even from spurious reflections. The unreliable candidates are filtered out by the agreement test and rank

the rest candidates by their reliability to find accurate target position. We can estimate the receiver's position even in the condition of attenuated direct path signal or high reverberation by using the proposed method.

Third, we proposed a low-complexity localization method to work in the highly reverberant environment. This method is based on the particle filter that estimates the position by weighted particles whose weights are computed by the likelihood. We designed likelihood function that efficiently calculates likelihood in the region with the direct path signal so that more reliable position can be obtained. The proposed method enables location estimation with high precision with a relatively small amount of computation in severe reverberation.

The proposed methods are evaluated in simulated environments with different reverberation time. The performances are verified in different parameters and compared with other localization methods. In addition, the performance is evaluated in the real reverberant environment with a large space. A series of experiments has shown the superiority of the proposed methods and it is appropriate to apply in the actual environment.

Keywords: location based services (LBS), indoor localization, acoustic receiver localization, time delay estimation, reverberation, particle filter

Student number: 2010-20905

Contents

List of Figures

List of Tables

Chapter 1

Introduction

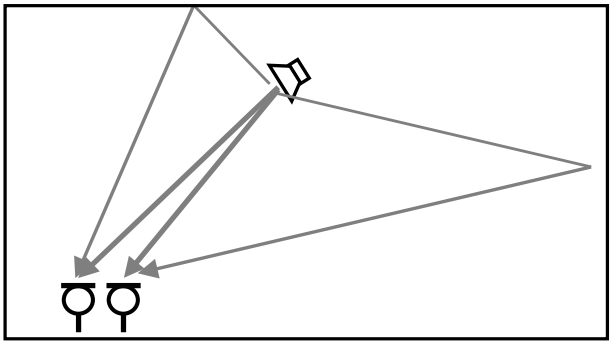
Location-based services (LBS) has become one of the essential parts in user-customized services due to the increasing use of mobile devices. Most of the widely used LBS are based on global positioning system (GPS) due to its high accuracy and easy access, but its operation is limited to the outdoor environments where the line of sight (LOS) is guaranteed. However, the demand for indoor LBS is rapidly increasing due to the growth of the virtual tour guide at museums, automated home, and internet of things (IoT) technology such as indoor assistant robots, tracking or control of home appliances.

One way to achieve the indoor LBS is to use the acoustic signals in the localization task. There are several advantages when acoustic signals are employed. Acoustic signals can be detected through obstacles even in unknown environments while other methods (e.g., RFID, ultrasonic, Bluetooth, or Wifi) are rather sensitive to LOS condition or prior knowledge about the environments [1], [2]. Moreover, it is efficient to use acoustic signals since there is no need for additional communication devices such as sensors or access points. Loudspeakers are pre-installed in most of

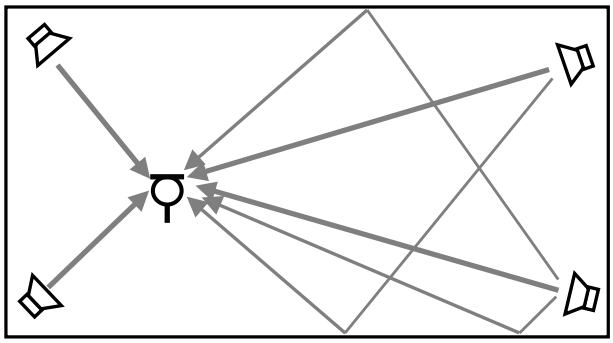
the buildings for reproducing voice announcements or playing background music and mobile devices such as cellphones or tablets are already equipped with microphones and loudspeakers.

The acoustic localization is studied in three ways: sound source localization (SSL), acoustic receiver localization (ARL) and self-localization. SSL aims to find the location or direction of the target source using multiple microphones placed at known positions. An example of the SSL case is shown in the Fig. 1.1a. It is often studied to improve the performance of speech recognition, speech enhancement or acoustic event detection [3]–[6]. The main goal of ARL is to find the target microphone position using the reference to the location of sound sources [7]–[12]. ARL is a study of finding the target microphone position with reference to the sound source positions. An example of the ARL case is shown in the Fig. 1.1b. The sound sources are assumed to be located at fixed known position and they are often pre-designed to meet the purpose of positioning. Self-localization is a research that estimates the location of each node in a wireless acoustic sensor network (WASN) where many microphones and loudspeakers are scattered [13]–[17]. An example of the self-localization case is shown in the Fig. 1.1c.

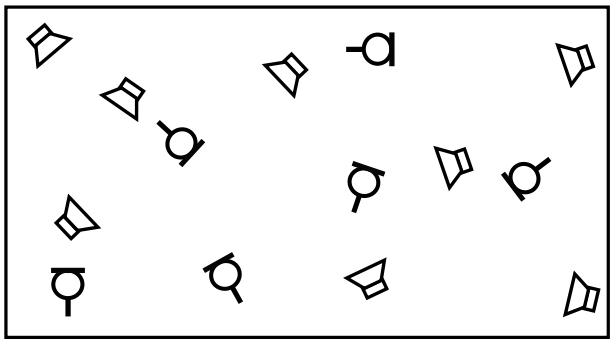
In a typical indoor localization scenario, the acoustic sensors (e.g., microphones) are easy to be deployed and loudspeakers are usually pre-installed or easy to be controlled. Since the main focus of indoor localization in an IoT environment is to estimate the relative position of home appliances or users, ARL can be considered as an optimal solution for such task. Moreover, the usage of ARL can minimize the resource requirement for acoustic indoor localization because it is fully operational only with the single microphone, whereas SSL and self-localization usually require multiple or array of microphones.



(a) An example of acoustic source localization.



(b) An example of acoustic receiver localization.



(c) An example of acoustic self-localization.

Figure 1.1. Cases of acoustic localization systems.

In an ARL system, the position is usually calculated from the time of arrival (TOA) or time difference of arrival (TDOA) which is estimated from the time delay of direct path signal. However, in highly reverberant environment, time delay estimation (TDE) is difficult because some reflections are often stronger than the direct path signal [18], [19]. Previous studies attempted to suppress reverberation by surrounding the environments by curtains [9] or conducting experiments in a small space where direct path signals can be easily identified [7], [12]. There are a few studies that attempt to tackle the multipath effect by identifying the direct path signal in reverberant environments [10], [20]. Studies in the SSL also focused on spurious measurements in the reverberant environments. DATEMM [21] tried to disambiguate TDOA measurements in the multiple source localization system and Canclini *et al.* [22] suggested a reliability criterion to disambiguate TDOA measurements in the reverberant condition.

In this thesis, we aim to estimate the position of the target user with an acoustic sensor based on the recording of signals from the fixed loudspeakers installed around the room. We target to estimate the position of the acoustic sensor with high accuracy and low-complexity in a large space with high reverberation. To do this, we present various localization techniques for the ARL system. We verify each technique in the simulated and the real environments.

The Chapter 2 explains the overall process of position estimation using acoustic signals. We first explain the types of position estimation using acoustic signals and describe the configurations of our system. For the transmitting part of the system, we explain the structure of the source signals previous studied in the ARL. For the receiving part, the process of time delay estimation will be described. Then the methods of position estimation from the time delays will be introduced. In addition,

we will discuss the practical problems that need to be considered in order to operate in real environments.

In the Chapter 3, we propose the source data structure to operate in the target environment. It is designed to distinguish a number of sources in the inaudible frequency band for the practical use in a real environment. Also, we propose a method to estimate the time delay of the direct path signal to reduce position estimation errors in the reverberant environments. The proposed method is compared with other methods and its performance is shown through simulation and actual recordings.

In the Chapter 4, we propose a method for accurate location estimation in highly reverberant environments. In the high reverberation condition, more spurious reflections occur, which makes it difficult to estimate the time delay of the direct path signal. If the time delay estimation is wrong, it is likely that the position estimate does not converge by an estimation method. In the proposed method, position candidates are obtained from most of the received signals including signals even from spurious reflections and the position is estimated by investigating multiple sources jointly. The unreliable candidates are filtered out by the agreement test and rank the rest candidates by their reliability to find accurate target position. The performance of the proposed method is evaluated in simulation and actual recordings.

In the Chapter 5, the ARL system is based on the particle filter. It estimates position by the weighted particles whose weights are computed from the likelihoods. We designed likelihood function that efficiently calculates likelihood in the region with the direct path signal so that more reliable position can be obtained. The proposed method enables location estimation with high precision with a relatively small amount of computation in severe reverberation. The performance of the proposed method is shown by simulation and actual recordings.

The rest of the thesis is organized as follows: The next chapter introduces the overall process of acoustic indoor localization. In Chapter 3, source data structure and localization method for reverberant environment are proposed. In Chapter 4, an accurate localization algorithm for the highly reverberant environment is proposed. In Chapter 5, the particle filter based positioning method with high precision with low-complexity is proposed. The conclusions are drawn in Chapter 6.

Chapter 2

Acoustic Receiver Localization System

In this section, we explain the task of indoor acoustic receiver localization using artificial signals. The overall idea of our system is depicted in Fig. 2.1. We target a large reverberant room where we need to consider the attenuation of acoustic signals and interferences of reflections. The loudspeakers are placed around the room and the target single-channel microphone is located inside it. The goal is to find the position of the receiver by the recording of acoustic signals played from loudspeakers. We will explain the structure of the source data and processing of the received data for localization in the following subsections. In addition, we will explain the practical issues in the application of the ARL system to the real environments.

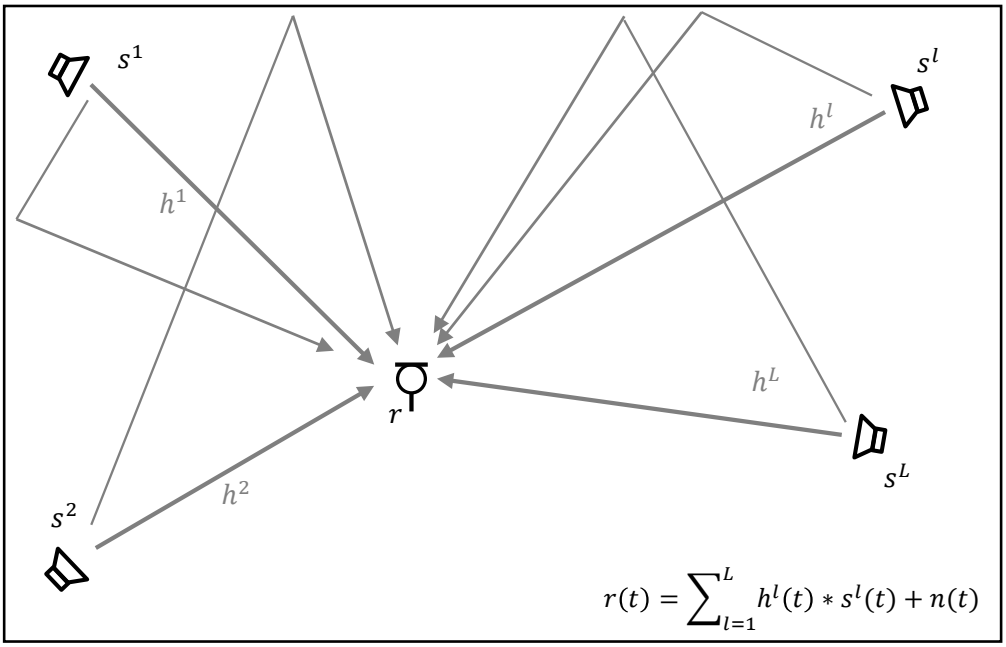
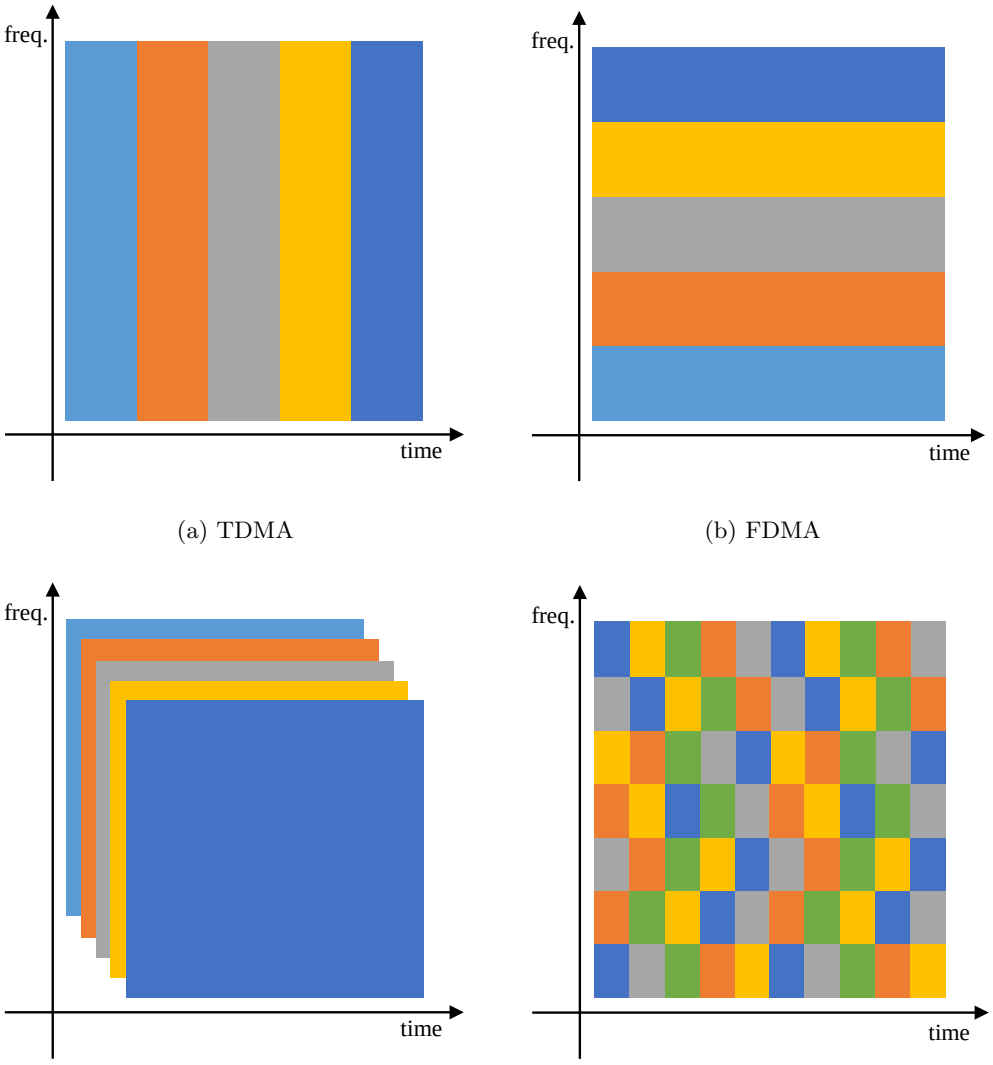


Figure 2.1. Relation and arrangement of the receiver and the sources. $s^l(t)$, $r(t)$, $h^l(t)$ and $n(t)$ represent the l-th source, the received signal, their RIR and background noise, respectively.

2.1 Source data structure

Since the received signal $r(t)$ is given as a superposition of a number of reverberated sources and noise as in the Fig. 2.1, the source signal $s^l(t)$ should be designed carefully in order to provide high accuracy in estimating time delays. In the ARL system, various source data structures have been studied. Since the ARL system can be thought of as a type of communication system, ideas of reference signal design are borrowed from the field of wireless communications. In order to distinguish multiple sources, multiple access schemes [23]–[26] such as time division multiple access (TDMA), frequency division multiple access (FDMA), code division multiple



(a) TDMA

(b) FDMA

(c) CDMA

(d) OFDMA

Figure 2.2. Schematics of widely used multiple access schemes: TDMA, FDMA, CDMA and OFDMA.

access (CDMA), or orthogonal frequency division multiple access (OFDMA) are applied and modified in ARL systems.

TDMA is a multiple access scheme that allocates specific time slots to each user using all the frequency bands indiscriminately, which is depicted in the Fig. 2.2a. When the number of users increases, the time slots are allocated in the same manner as the round robin method. Since all the frequency bands are used, it is robust to noise and has low data losses. The process of synchronization is required in order to communicate in the assigned time slot. Some previous studies employed TDMA as the source data structure of acoustic localization system to distinguish the identity of each loudspeaker [7], [10], [27]. A short signal is transmitted for a predetermined period of time in order, and a microphone uses the corresponding source signals to perform synchronization to distinguish each source.

FDMA is a multiple access method that assigns different frequency bands to each user, which is depicted in the Fig. 2.2b. In this method, data is transmitted and received only through pre-assigned frequency bands to each user. If the number of users increases, the frequency bandwidth may be reduced and interference may occur between adjacent bands. To reduce the interference, guard bands are often inserted between each band to separate them. Some studies with FDMA based acoustic communication system, frequency limited signals like sine waves are used as a method for determining the presence of a signal. The Doppler effect should be considered when using data structure dependent to frequency because the frequency of the signal is shifted as the source or the receiver is moved [24].

CDMA is a type of multiple access method that allows several users to share the same frequency band by assigning the special code to each user using spread-spectrum technology [23]–[25] and it is depicted in the Fig. 2.2c. This method is

widely used in mobile communications. However, since users share the frequency bands and time slots, it is necessary to consider the near-far effect. The near-far effect refers to a situation when the desired signal is far away, it is difficult to receive the desired signal due to the interference of closer unwanted signals. In acoustic localization system, due to its simple use and implementation, many previous studies [7], [9], [11], [12], [27] are based on the CDMA. However, due to its issues of application to real environments, the experiments are usually conducted in the controlled environment where it is easy to detect the direct signals.

OFDMA is the multiple access technique to accommodate users in a given bandwidth. It is based on the modulation method called orthogonal frequency division multiplexing (OFDM) [23]–[25]. OFDM divides a channel into multiple orthogonal bands so that they don't interfere with each other. OFDMA assigns a group of subcarriers to each user to achieve multiple access and it is often combined with TDMA, FDMA, or CDMA.

There are studies using other multiple access methods like chirp signals as the source data structure of the ARL system [28]–[31]. The chirp signal is the signals with slant frequency which are usually in inaudible frequency bands. It distinguishes each signal by modifying the slope of frequency or assigns a specific code to the frequency domain and makes codewords.Some studies employed the chirp signal and used it as the spread spectrum [29].

In ARL system, it is important to make a proper structure to distinguish each source from the mixture. In the process, the source data structure should be designed according to the purpose of the system (e.g., locating or tracking), or the target environment (e.g., noisy environment or band-limited environment). Although it is possible to refer to advantages and disadvantages of various multiple access used in

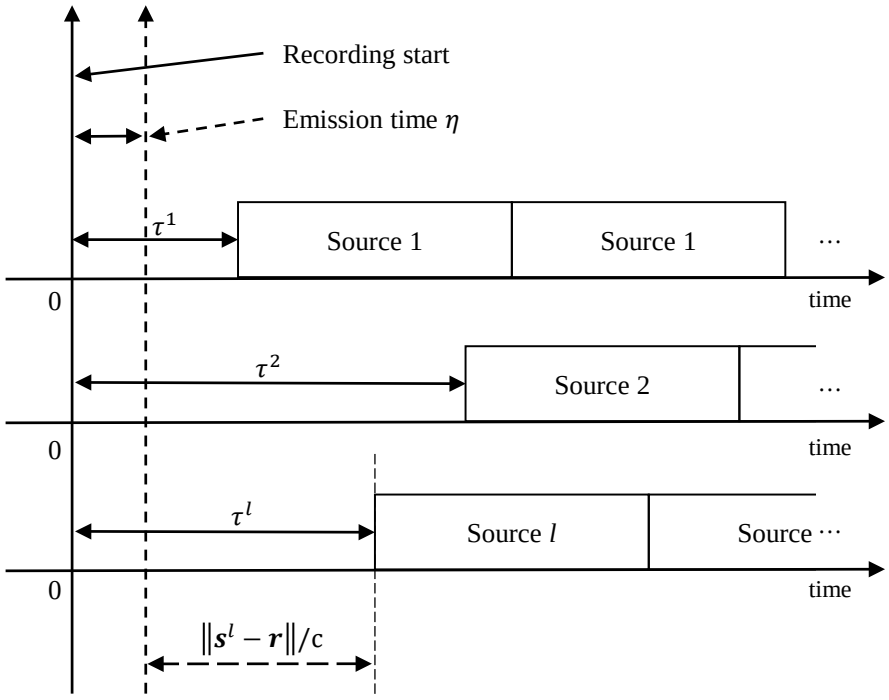


Figure 2.3. The relation between the time delays and the emission time. The sources are played concurrently at an unknown emission time. The time delay is the sum of the emission time and the delay from the distance between the sources and receiver.

communication systems when designing the source data structure, it should be noted that the acoustic data communication is a passive system, that is, the transmitting and receiving ends are separated.

2.2 Localization from the received signal

In this section, we explain the process of localization from the received signal: distinction of the each source, TDE and localization. The target receiver is assumed to be enclosed by a set of loudspeakers placed in a reverberant room as shown in

Fig. 2.1 and records all of the incoming source signals with the single-channel microphone. The relation of the direct propagation time t^l_{prop} and the positions of the source $\mathbf{s}^l$ and receiver $\mathbf{r}$ can be represented as

$$t^l_{prop} = \frac{\parallel \mathbf{s}^l - \mathbf{r} \parallel}{c} + \eta \tag{2.1}$$

where superscript l, η and c indicate a source index, the source emission time and the speed of the sound, respectively and $\parallel . \parallel$ represents the Euclidean norm. This relation is depicted in Fig. 2.3. The emission time η is an unknown variable if the source and the receiver are not synchronized. In case of unsynchronized system, it is impossible to estimate the position if η is unknown or estimated with wrong value [13], [15], [32].

In the TOA-based approach the time delay t^l can be straightforwardly converted to pair-wise distance d^l using the relation $d^l = ct^l$. Though simple to implement, the TOA-based approach requires additional process of synchronization between the receiver and the sources to obtain η [8], [9], [17].

In the TDOA-based system, the receiver and the sources are assumed to be asynchronous hence the time delays $\{t^l\}$ cannot be directly used to calculate the distance. Instead, the position of the receiver is estimated from the time differences $\{t^{lm}\}$. Since the TDOA is calculated by the difference of TOAs as

$$t^{lm}_{prop} = t^l_{prop} - t^m_{prop} \tag{2.2}$$

$$= \frac{\parallel \mathbf{s}^l - \mathbf{r} \parallel}{c} + \eta - (\frac{\parallel \mathbf{s}^m - \mathbf{r} \parallel}{c} + \eta) \tag{2.3}$$

$$= \frac{1}{c}(\parallel \mathbf{s}^l - \mathbf{r} \parallel - \parallel \mathbf{s}^m - \mathbf{r} \parallel), \tag{2.4}$$

consideration of η can be omitted because we assume that all the sources emit signals simultaneously. In this paper, we assume the case of unknown η since it is

more practical. However, because the TDOA is a function of the maximum peak indices of sources, i.e. t^l_{max}, it is difficult to deal with spurious peaks caused by multipath in the reverberant condition.

We now model the above localization system to estimate time delays $\{t^l\}$ as following. The room impulse response (RIR) can be represented as a summation of weighted Dirac delta functions $\delta(t)$ as following:

$$h^l(t) = \sum_j a^l_j \delta(t - \tau^l_j) \tag{2.5}$$

where a^l_j indicates the amplitude of the j-th multipath of the RIR at time τ^l_j. Based on (2.5), the received signal $r(t)$ at time t can be represented as

$$r(t) = \sum_{l=1}^{L} h^l(t) * s^l(t) + n(t) \tag{2.6}$$

where $*$, $s^l(t)$ and $n(t)$ indicate convolution, l-th source signal and the background noise, respectively.

To compute the position from $r(t)$, it is important to estimate the time delay t^l between the receiver and each source as shown in Fig. 2.5. To estimate t^l, we need to find the peaks in the cross-correlation between $r(t)$ and $s^l(t)$. The generalized cross correlation (GCC) [6], [33] is widely used method in computing cross correlation which can be obtained by

$$R^l(t) = \mathcal{F}^{-1}[\mathcal{F}[s^l(t)]\mathcal{F}^*[r(t)]], \tag{2.7}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and its inverse transform, respectively and $\mathcal{F}^*$ represents the complex conjugate of $\mathcal{F}$. The time delay t^l is estimated by the index of the maximum peak of the $R^l(t)$ which can be written as

$$t^l = \underset{t}{\operatorname{argmax}} \, |R^l(t)| \tag{2.8}$$

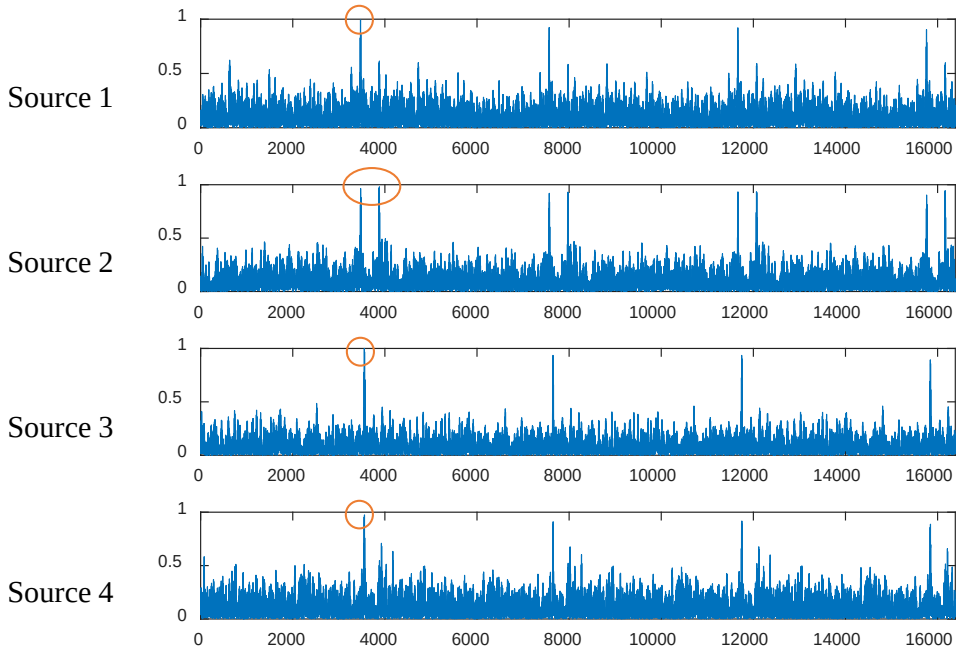


Figure 2.4. An example of cross-correlation of each source with the received signal. The time delays are obtained from the peaks of each cross-correlation.

where $|.|$ represents the absolute value. There will be a train of peaks with a cycle of the frame length since our system plays sources repetitively. An example of $\{R^l(t)\}$ is shown in the Fig. 2.4. For accurate localization, it is important to estimate the time delay of the direct propagation which is $t^l = t^l_{prop}$.

After estimating $\{t^l\}$ and η, the final position estimation can be performed in various methods. The most commonly used method is to solve nonlinear equations, for example,

$$\tilde{\mathbf{r}} = \underset{\mathbf{r}}{argmin}(c\tilde{t^l} - \parallel \mathbf{s}^l - \mathbf{r} \parallel)^2 \tag{2.9}$$

where $\tilde{t^l}$ is the observed time delay and $\tilde{\mathbf{r}}$ is the estimated position. An algorithm such as the Newton method or Levenberg-Marquardt algorithm (LMA) is used to

solve this kind of equations [34]–[37]. Although it is a simple computation, if there are errors in TDE, i.e. TOA or TDOA, the algorithm will not converge and the answer will end up in a wrong location. The proposed methods in Chapters 3 and 4 use LMA for the position estimation method [37].

Another position estimation technique is the particle filter. Since it calculates the position by weights of particles, it is often unnecessary to specify estimated time delay for each source. Even with the inaccurate TDE, it is likely to provide a similar location in the vicinity if we properly design the likelihood function. For this reason, many localization studies [38]–[40] employ the particle filter approach. Since we target environment with high reverberation, it is probable to face ambiguous time delays. We employ the particle filtering as the localization method in Chapter 5.

2.3 TDE in reverberant environments

For a practical application of TDE to indoor localization, it is essential to consider the effect of acoustic reverberation. Time delays such as TDOA or TOA are estimated by investigating time lags on the correlation between the original sources and received signals. In an ideal case, the time lag is decided by picking the peak where the correlation is maximized. In the presence of reverberation, however, the peak detection may not be reliable due to reflective paths which can be almost always seen in usual RIRs.

The early reflections are the part of an RIR that appears right after the direct sound as shown in Fig. 2.5. The early reflections consist of large impulsive signals up to 50 ms of the RIR in the time domain, which is caused by the nearby walls

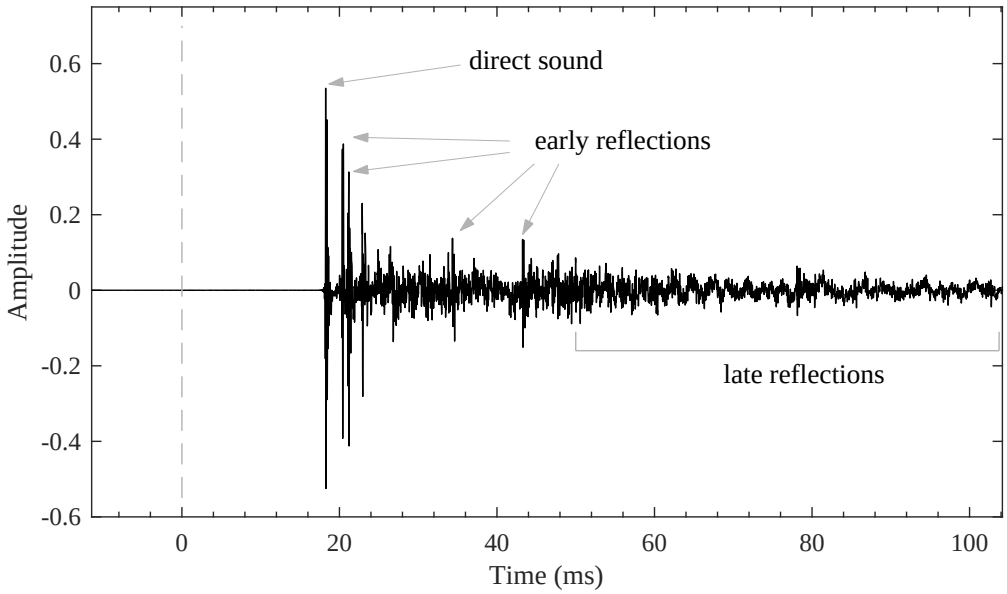


Figure 2.5. An illustration of the direct sound, early reflections and late reflections in the RIR between the receiver and the source.

or objects [41]–[43]. These early reflections incur major ambiguity during the peak detection process because the magnitude of the peaks caused by the early reflections is sometimes comparable to or even larger than that of the direct path signal. For this reason, it is necessary to carefully examine the preceding parts of peak correlation when estimating the position in reverberant environments [18]–[20].

The late reflections and the attenuation of the source signal also make it difficult to retrieve the desired peak. The signals arriving after the early reflections are called the late reflections which can be seen in Fig. 2.5. They usually seem like randomly distributed decaying signals in the RIR, which act as noise in severe reverberation [41]–[43]. If the source signal is generated far from the receiver, it will be attenuated according to the inverse square of the distance. As a result, it will easily be masked by nearby strong signals, late reflections or noise making it difficult to retrieve the

target signal, which is called as the near-far effect.

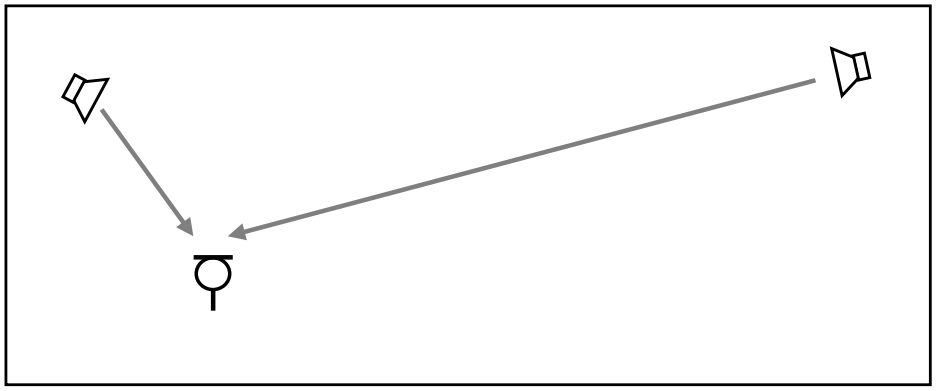
2.4 Near-far effect

The position estimation in the ARL system is done by the communication between the acoustic transmitter (loudspeakers) and the acoustic sensors (microphones). The near-far effect is one of the major problems in the wireless communication systems [25]. Since our ARL system is a kind of communication system, it is necessary to consider the near-far effect.

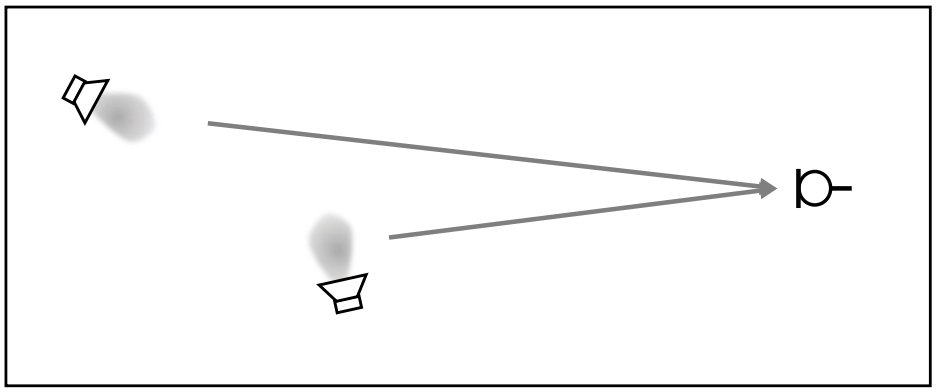
The near-far effect means that the desired signal is difficult to detect due to the interference of the closer signal. The example of the near-far effect is shown in the Fig. 2.6a. Also, the beam pattern of the source and the receiver can cause the near-far effect as shown in the Fig. 2.6b. That is, the near-far effect occurs in situations when the desired source and the interfering sources share the same frequency bands and time slots [25], [44].

There are many studies to overcome this effect in the field of the wireless communications. One of the general solutions in wireless communications is the power control technique. In wireless communication, the system can amplify the power of the desired signal so that it becomes strong enough to be retrieved regardless of the interference. Unfortunately, however, in ARL there is no way to feedback the channel information supporting only one-way communication so it is impractical to apply the power control technique.

An alternative way to tackle the near-far effects is to design the source signals such that they reside in non-overlapping frequency regions. We will introduce the solution to this problem by designing the source data structure in the next chapter.



(a) The case of the near-far caused by the closer signal interfering the desired signal from a far distance.



(b) The case of the near-far caused by the beam pattern of the loudspeakers. A similar situation can happen due to the beam pattern of the microphon as well.

Figure 2.6. Examples of the near-far effect.

Chapter 3

Indoor Localization using Inaudible Acoustic Signals

3.1 Introduction

One of the important issues in ARL is how to design the source signals, and in traditional approaches, specific sequences such as the pseudo-random sequences [8] or Gold codes [7], [9] are utilized, which are then modulated by means of the code division multiple access (CDMA) [7], [12], or direct sequence spread spectrum techniques [9], [29]. The location of the target acoustic sensor is usually estimated based on the time of arrival (TOA) or time difference of arrival (TDOA). In a TOA-based system, localization is performed by calculating distances between the receiver and the sources while considering the relation between the time delay and the speed of sound in a synchronous manner. In contrast, the TDOA-based methods allow asynchrony between the receiver and the sources and compute the position through multilateration or iterative estimation [28], [29].

Although previous studies on ARL have focused on accurate localization, they lacked in adaptation to actual environments. In real environments, there usually exist acoustic reverberation and the near-far effects implying the situation that the target signal is masked by nearby strong signals and makes it difficult to retrieve the target signal. Previous studies have conducted experiments in a small space surrounded by curtains [9] or extracted fingerprint [7] of the received signals to avoid reverberation or the near-far effects.

In this chapter, we focus on two principles for a practical application of the ARL. First, we aim not to influence on the human hearing by using inaudible acoustic signals which can be processed in off-the-shelf audio devices. Second, the localization algorithm should perform well in large reverberant room environments with sub-meter accuracy. Based on these principles, we propose an indoor localization system for which an efficient source data structure is designed to deal with environmental issues, and a direct path detection algorithm is developed to cope with the effects of multipath. Basically, localization is carried out by means of the time difference of arrival (TDOA)-based technique. The performance of the proposed algorithm is evaluated in both actual and simulated room environments.

3.2 Acoustic source design and synchronization

In this section, we propose a source data structure and synchronization algorithm for an efficient localization in reverberant environments. For this, we concentrate on several practical issues. First, the acoustic source signals should be designed to be robust to the reverberations which cause multipath effects. Second, some special techniques should be devised to mitigate the near-far effects. Third, we also need

to consider the case in which the power of indirect (multipath) signal component becomes stronger than the direct signal component.

3.2.1 Reverberation in multipath environments

The root mean square (RMS) delay spread is one of the important measures in understanding a multipath channel. The RMS delay spread σ_{rms} is defined as the standard deviation of the power of the room impulse response $h(t)$ in (2.5), as given by

$$\sigma_{rms} = \sqrt{\frac{\sum_j a_j^2(\tau_j - \bar{\beta})^2}{\sum_j a_j^2}} \tag{3.1}$$

where

$$\bar{\beta} = \frac{\sum_j a_j^2 \tau_j}{\sum_j a_j^2}. \tag{3.2}$$

For simplicity, we omit the super-script l which denotes the source index. If σ_{rms} is relatively short compared to the symbol duration, the intersymbol interference is prevented, i.e., the channel can be considered flat [23], [24]. We can consider the data frame of the source signal plays the similar role of the data symbol in the communication systems. For this reason, the length of the data frame should be determined long enough to deal with the multipath effects.

3.2.2 Source data structure for ARL

To make ARL work in the large space, the source data structure should be well designed to deal with the near-far effects. The near-far effects usually occur in situations when the desired source and the interfering sources share the same frequency bands and time slots [44]. In wireless communications, the system amplifies the

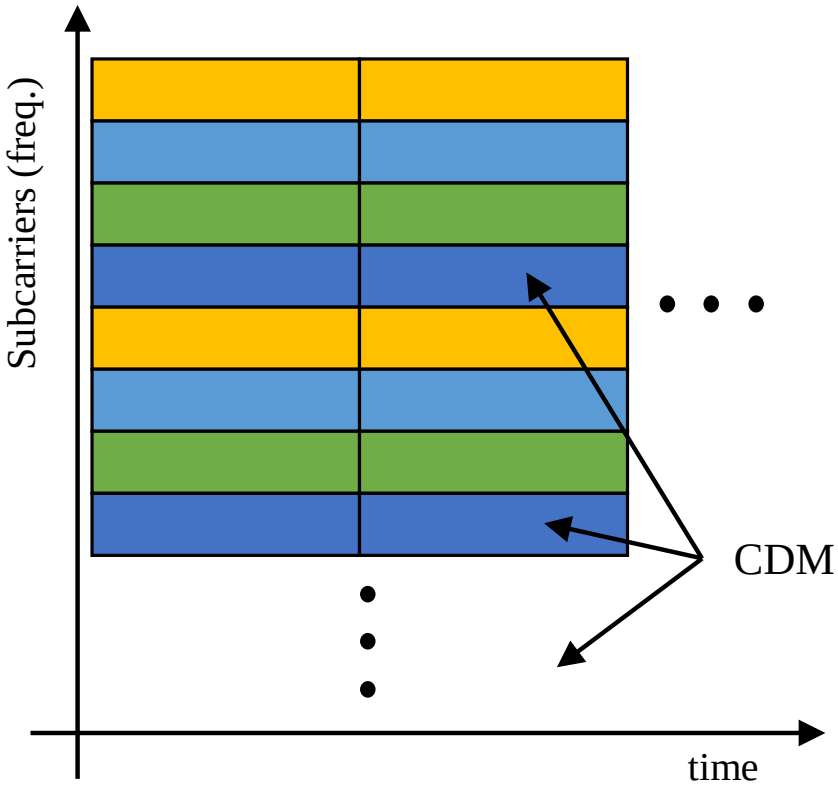


Figure 3.1. A schematic of the structure of OFDMA-CDM.

power of the desired signal so that it becomes strong enough to be retrieved regardless of the interference, which is called the power control technique. Unfortunately, however, in ARL there is no way to feedback the channel information supporting only one-way communication so it is impractical to apply the power control technique. An alternative way to tackle the near-far effects is to design the source signals such that they reside in non-overlapping frequency regions.

In order to achieve this, we borrow an idea from the orthogonal frequency division multiple access-code division multiplexing (OFDMA-CDM) approach [45] for the acoustic data structure design. OFDMA is a multiple access scheme that divides

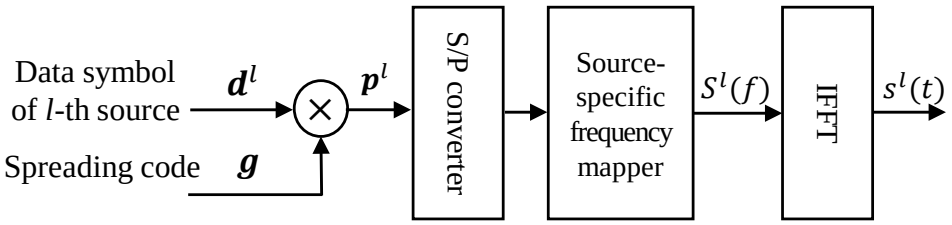


Figure 3.2. Schematics of generating the source data. Source data is generated from l-th data symbol $\boldsymbol{d}^l$ spread by a unique sequence $\boldsymbol{g}$ which are then followed by source-specific frequency mapping and inverse Fourier transform to time domain signal.

the bandwidth into closely spaced multiple subcarriers assigned to each source [23]–[25]. OFDMA-CDM employs OFDMA for multiple access and additionally applies CDM for the diversity of each source [23], [45], [46]. An idea of the structure of OFDMA-CDM is depicted in the Fig. 3.1.

In designing the sources, the characteristics of the cross-correlation among different sources and auto-correlation of each source are important. The auto-correlation function of each source needs to have a salient peak for a successful estimation of the time delay t^l. In contrast, the cross-correlation between different sources should be kept as small as possible to avoid interferences. Among many kinds of pseudo-random sequences, we apply the Gold sequence due to its good correlation properties [7], [9], [23].

Fig. 3.2 shows how the proposed algorithm generates the l-th source signal. The vector $\boldsymbol{d}^l$ represents the data symbol assigned to the l-th source. Each data symbol $\boldsymbol{d}^l$ is spread by a Gold sequence $\boldsymbol{g}$ resulting in the transmission vector $\boldsymbol{p}^l$. Then, $\boldsymbol{p}^l$ is interleaved evenly onto its subcarriers by the source-specific frequency mapper producing the source symbol $S^l(f)$. The source symbols $S^l(f)$ described

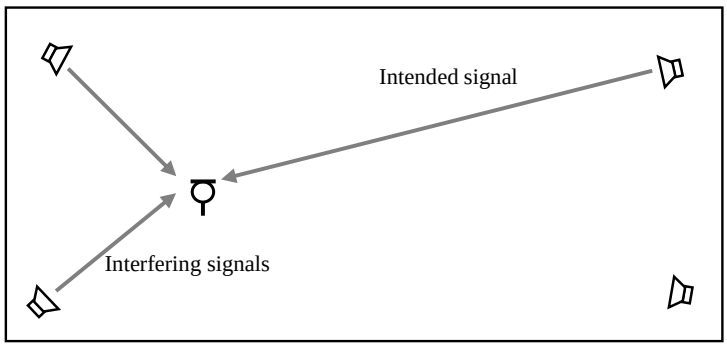
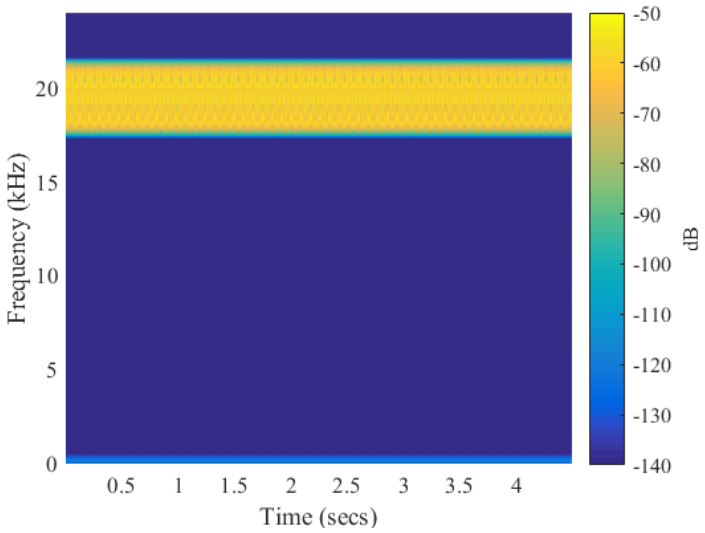


Figure 3.3. It is called as the near-far effect when the intended signal is interfered by the closer signals.

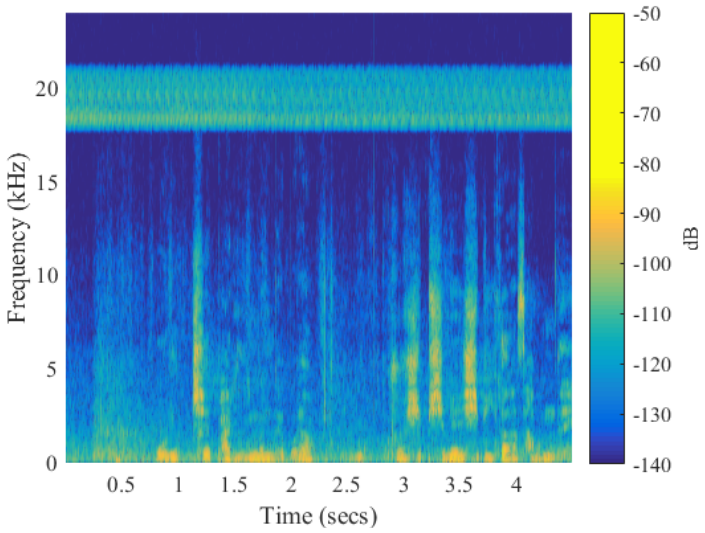
in the frequency domain is then transformed into the corresponding time domain source $s^l(t)$ through inverse FFT. Interested readers are encouraged to refer to [23], [45], [46] for the basic processing in OFDMA-CDM.

Here, the source-specific frequency mapper assigns each source to a non-overlapping subcarrier in a specified order, which has the following advantages: First, it enables to avoid the near-far effects since each source resides in non-overlapping frequency regions. Second, each source signal becomes robust to non-flat channel responses due to the combined effect of interleaving and diversity. This is very useful because the frequency responses of the commonly used microphones and loudspeakers cannot be regarded flat.

An example of spectrogram of the source signal is shown in Fig. 3.4a. We designed the signal to be band limited in inaudible frequency to affect the least to the human hearing by assigning subcarriers of high frequency to the frequency mapper. As can be seen in Fig. 3.4b, the received signal is attenuated and mixed with background noise as explained in (2.6). However, due to the frequency difference, the ordinary conversation or noise hardly affects the source signal.

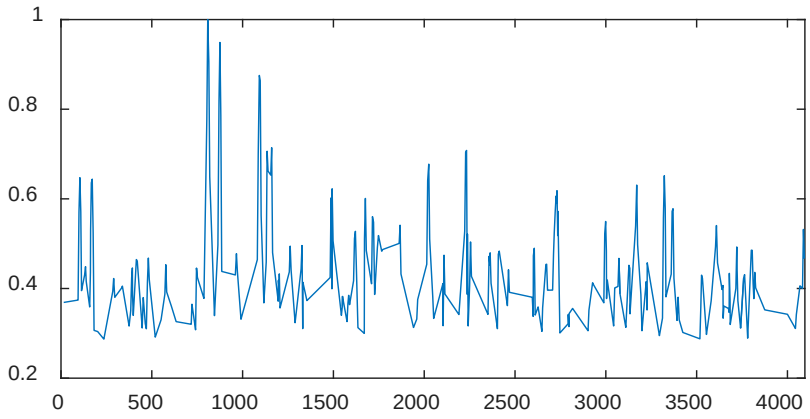


(a) Spectrogram of the source signal. The proposed source signal is bandlimited in inaudible frequency band.

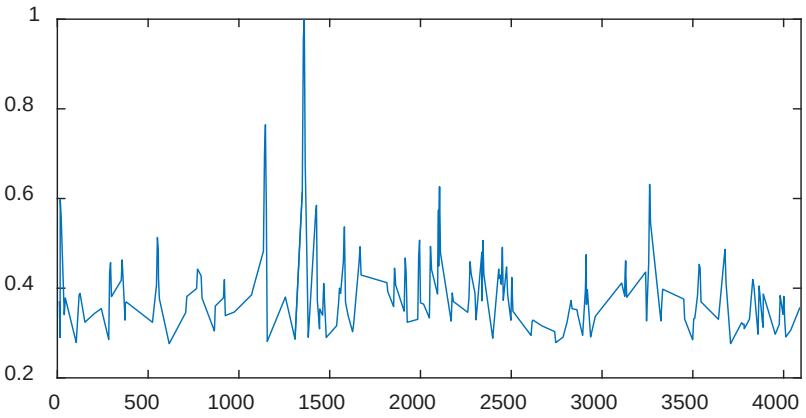


(b) Spectrogram of the received signal. The ordinary conversation does not affect the source signal.

Figure 3.4. Examples of spectrograms of the source signal and the received signal.



(a) The case of the maximum peak is the direct path siganl.



(b) The case when the direct path signal precedes the maximum peak.

Figure 3.5. Examples of cross-correlation between the source and the received signal.

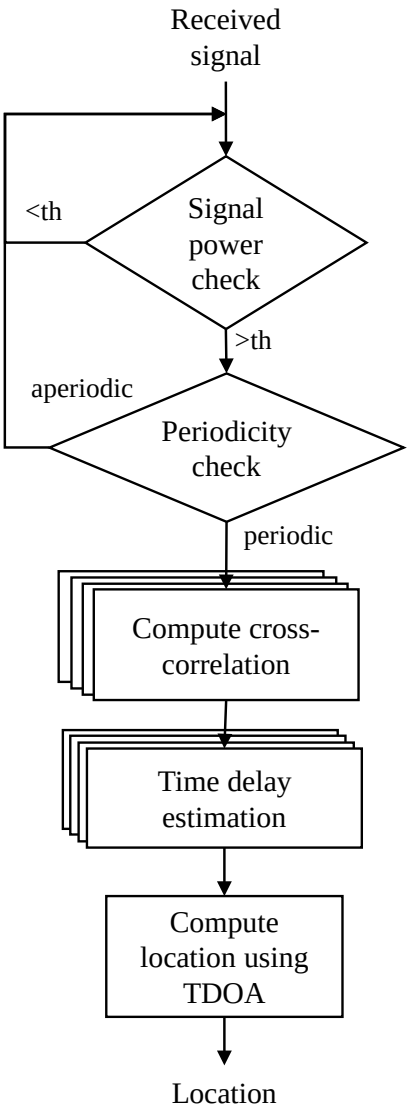


Figure 3.6. A schematic of the process of signal presence detection.

3.2.3 Signal presence detection

When considering practical use, it is necessary to check the presence of the desired signal in the received signal. In this section, we introduce a process of detecting the existence of a signal efficiently considering the characteristics of the source signal.

The characteristics of the generated signal are as follows:

- Band limited signal: to affect the least on the human hearing, the proposed signal resides on the narrow inaudible frequency band.
- Periodic signal: source signal with short data frame size is played repeatedly.

The process of signal presence detection using the above characteristics is shown in Fig. 3.6. First, the signal power in the predetermined band which is known in advance is computed and compared with a certain threshold. In everyday life, there is little signal in the inaudible frequency band, but in the case of white noise or acoustic events with impulsive noise may affect the inaudible frequency band. The second step checks the periodicity of the received signal to eliminate the noise cases. The cycle of the source signal is predetermined by the length of the data frame, the periodicity of the auto-correlation of the received signal is compared with the data frame length. If the received signal has the desired periodicity, it will proceed to TDE process. The presented two stage processes can efficiently determine the presence of the desired signal with low computational complexity.

3.2.4 Direct path detection

Another important issue is that the power of the first reflection component of a source signal sometimes becomes comparable to or even stronger than that of the direct path

component in real multipath environments. An example of the case is depicted in the Fig. 3.5b. Human auditory perception treats the preceding signal component more importantly when perceiving direction, which is called the precedence effect [18], [41], [47]. Conventional ARL systems mostly focused on finding the maximum peak in the cross-correlation. In ideal case with no or small reverberation, the maximum peak will be the direct path signal. However, in reverberant environments, the shape of the cross-correlation is distorted by the early and late reverberation. Even though the maximum peak indicates the direct path signal for most of the time as in the Fig. 3.5a, the case where the direct path signal precedes the maximum peak as in the Fig 3.5b will lead to error in position estimation. Some studies introduce parameters to emphasize peaks in the cross-correlation by peak-quality features but they only enhance the highest peak [7], [8]. Only a few studies have attempted to consider the precedence effect [19]. Based on this, it is important to consider not only the absolute level of the correlation peaks but also their precedences when estimating the TOA or TDOA.

We propose a simple but effective method which refines the result of the synchronization process given in (2.7) and (2.8). In this method we derive the modified peak time indices $\{t^l_{prec}\}$ in the following way:

$$t^l_{prec} = \begin{cases} \hat{t}^l, & \text{if } R^l(\hat{t}^l) > \xi R^l(t^l_{max}), \\ t^l_{max}, & \text{otherwise,} \end{cases} \tag{3.3}$$

where

$$\hat{t}^l = \underset{t^l_{max}-W_{prec}+1 \le t < t^l_{max}}{\operatorname{argmax}} R^l(t). \tag{3.4}$$

The highest peak time index obtained from (2.8) is indicated by t^l_{max}, and W_{prec} and ξ denote the length of the searching window and the threshold ratio of the peak,

Table 3.1: Parameters of the real environment

Parameter	**Value**
Dimension (m)	7.99 × 10.36 × 3.90
Loudspeaker position (m) (height: 0.94 m)	(0.87, 1.11), (6.77, 1.31), (6.39, 8.79), (1.68, 8.92)
RT_{60} (seconds)	0.571
RMS delay spread (seconds)	0.032
Number of target positions	60 (C1000S), 52 (iPad3, GalaxyS3)

Table 3.2: Simulated environment

Parameter	**Value**
Dimension (m)	8 × 10
Loudspeaker position (m)	(1,1), (7,1), (1,9), (7,9)
RT_{60} (seconds)	0.5, 1.1
Number of target positions	165

respectively. In our experiments, W_{prec} and ξ were determined empirically based on the room acoustics [18], [41].

3.3 Performance evaluation

Experiments were conducted to evaluate the performance of the proposed approach to ARL. As a performance measure, the localization error is calculated by the root mean square error (RMSE) of the distances between the measured and estimated

Table 3.3: System configuration

Parameter	**Value**
Sampling rate (kHz)	48
Frequency band (kHz)	18 - 21.5
Frame length (samples)	4096, 8192
W_{prec} (samples)	500
ξ	0.65

positions over whole data frames for each target location. The result is depicted as the cumulative distribution of the localization error.

3.3.1 Experimental setup and system configuration

We conducted experiments in a large reverberant seminar room and the simulated environments. Four loudspeakers were placed around the surrounding walls. For the simulation of reverberation, the RIRs corresponding to RT_{60} of 0.5 and 1.1 seconds were generated by Allen and Berkley's image method [48]. Total 100 test positions were uniformly spread and RMSEs were calculated.

For the experiment in the real environment, Genelec 8030 loudspeakers and AKG C1000S microphone were used for this experiment. The environmental parameters of the room are listed in Table 3.1 and those of the simulated room in the Table 3.2. The reverberation time (RT_{60}) was measured using pink noise in a preliminary experiment [49], and σ_{rms} was calculated from the measured RIR.

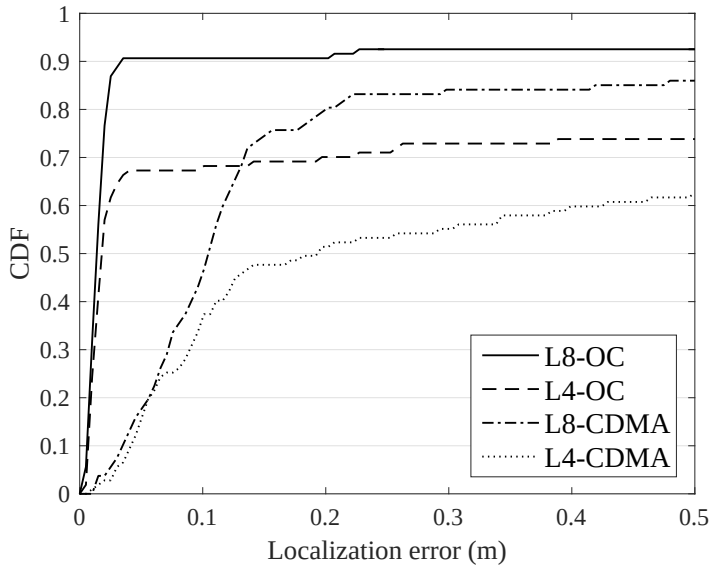
Acoustic source signals were generated in the frequency bands above 18 kHz for inaudibility. The source signals are played concurrently and repeatedly. The

frame length of the source data was determined to be sufficiently longer than σ_{rms}. Synchronization parameters W_{prec} and ξ were set based on the room acoustics [18], [41]. The configuration of the ARL system constructed for performance evaluation is given in Table 3.3.

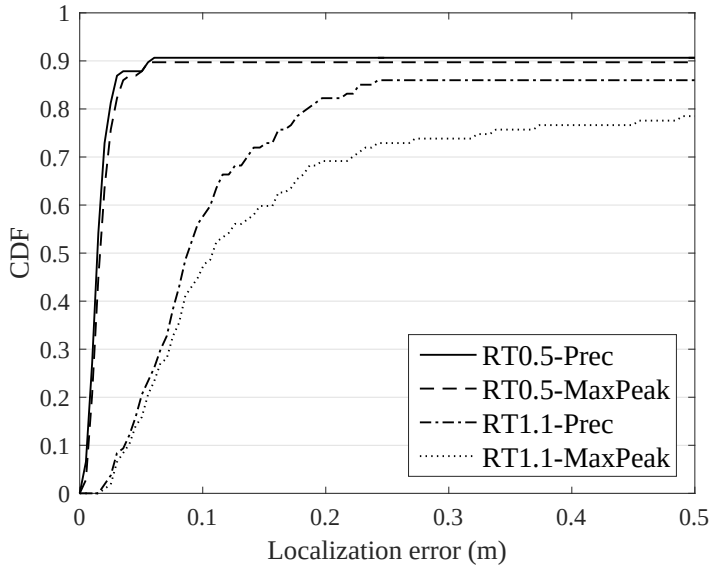
3.3.2 Evaluation of acoustic data structure

The first simulation was conducted to confirm the structure of the acoustic source data. The simulation result in Fig. 3.7a shows the comparison between the proposed OFDMA-CDM data scheme (OC) and a conventional CDMA-based scheme for different data frame lengths of 8192 ($L8$) and 4096 ($L4$) samples. The CDMA-based system was implemented as a conventional structure from [9] with some modification for a fair comparison. The RT60 was fixed to 0.5 seconds for the reverberant environment simulation and the average RMS delay spread calculated from the generated RIRs was 1120 samples. All the signals used in this experiment had the same sampling rate, frequency band and frame length as shown in Table 3.3.

We can see that longer data frame length cases ($L8$) are more robust than the shorter ones ($L4$). This is because the data frame length in this experiment was sufficiently longer than the RMS delay spread which the channel can be regarded flat [23], [24]. For the structural difference, the OC outperformed the CDMA-based source data structure. This can be accounted for by the fact that the CDMA-based technique separates the source data in the code domain resulting overlaps in the frequency domain, which is not suitable for mitigating the near-far effects.



(a) An experiment of the proposed source data structure (OC) in different frame length.



(b) Comparison of the proposed algorithm with the conventional method in different RT.

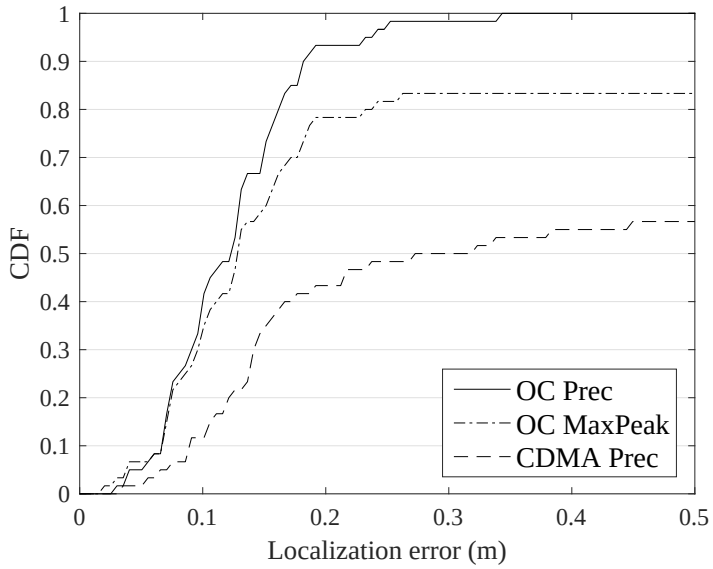
Figure 3.7. The performance of the proposed in simulated environments.

3.3.3 Performance of the direct path detection algorithm

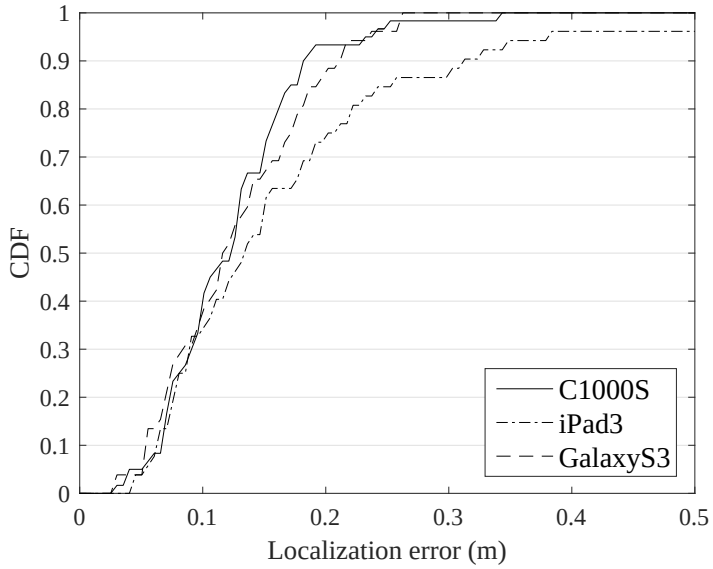
The second experiment was conducted to evaluate the effect of the direct path detection algorithm in different reverberant environments. Using the parameters of the previous experiment, the data frame length was fixed to 8192 samples (*L8*) and performance was measured by the *OC*. For simulation of severe reverberant environments, the RIR with RT60 of 0.5 and 1.1 seconds cases were generated and denoted as *RT0.5* and *RT1.1*, respectively. The case of direct path detection algorithm and the conventional synchronization method using the maximum peak as in (2.8) are denoted as *Prec* and *MaxPeak*, respectively. From the result shown in Fig. 3.7b, we can see that *Prec* outperformed *MaxPeak* in all environments. One notable observation from this experiment is that the performance of *Prec* was more effective than *MaxPeak* in severe reverberation (*RT1.1*). This result indicates that there exists more portion of stronger indirect signals in reverberant environments.

3.3.4 Performance in a real room

We compared the performances of three different schemes of ARL in the real environment. The first scheme is the ARL technique with the proposed source data structure and the direct path detection algorithm, which is denoted by *OC Prec*. The second scheme denoted by *OC MaxPeak* has the same source data structure to that of *OC Prec* but applies the conventional peak detection algorithm given in (2.8). The last scheme denoted by *CDMA* represents a conventional CDMA-based system presented in [9] with some modification for a fair comparison. All three schemes had the same sampling rate, frequency band and frame length as shown in Table 3.3. RMSEs were calculated at 60 different positions inside the room. The results



(a) The performance comparison with different settings in the actual reverberant room.



(b) The performance of the proposed method on different devices: C1000S, iPad3, and GalaxyS3.

Figure 3.8. The performance of the proposed in the real environment.

are shown in Fig. 3.8a where we can see that the proposed OFDMA-CDM data structure (*OC Prec* and *OC MaxPeak*) outperformed the CDMA-based source data structure. This can be accounted for by the fact that the CDMA-based technique separates the source data in the code domain resulting in an amount of overlap in the frequency domain, which is not suitable for mitigating the near-far effects. In addition, the superior performance of *OC Prec* to that of *OC MaxPeak* confirms the effectiveness of the proposed synchronization algorithm.

Next, we evaluated the performance of the proposed approach, i.e., *OC Prec* with different types of receiver devices. We computed RMSEs at 52 positions using two different receiver devices: Apple *iPad3* and Samsung *GalaxyS3*. The result is shown in Fig. 3.8b, where the result obtained with *C1000S* microphone is also displayed. One notable observation from this result is that the performance of *OC Prec* is almost similar with various off-the-shelf receiver devices.

3.4 Summary

This chapter has presented a novel indoor localization system using the inaudible acoustic signal that operates in real environments. We have proposed the structure of acoustic source data and synchronization algorithm for localization to deal with multipath in the reverberant environment and near-far effects. The proposed approach has shown good performance in both the actual and simulated room environments.

Chapter 4

Robust Time Delay Estimation for Acoustic Indoor Localization in Reverberant Environments

4.1 Introduction

In this chapter, we propose a robust time delay estimation (TDE) algorithm to cope with the reverberant environments in ARL. In the actual environments, it is difficult to estimate the correct time delays of the direct path signals for localization due to multipath signals caused by acoustic reverberation. Although previous studies on ARL focused on accurate localization, they lacked in adaptation to actual environments. Aloui et al. [7] performed ARL experiments in a small space where the direct signal was dominant. Sertatıl et al. [9] succeeded in measuring the receiver position within a small area surrounded by curtains. Some other studies attempted ARL in large spaces which are likely to be reverberant, but the reverberation was

not their major concern [8], [29]. Some of the recent works in SSL focused on ambiguous TDOA measurements caused by reverberation. DATEMM [21] suggested disambiguation of TDOA estimation for multiple sources using microphone arrays and the work is followed by [50]. Canclini et al. [22] performed TDOA disambiguation with the use of a reliability criterion in a distributed microphone array network system.

The proposed algorithm consists of two parts: minimizing the effect of multipath caused by reverberation and determining the receiver position via a reliability measure. We target this algorithm to perform well in large reverberant room environments. The performance of the proposed algorithm is evaluated with an asynchronous TDOA-based localization system in both actual and simulated room environments.

4.2 Robust TDE

In this section, we propose a robust TDE algorithm for application to practical reverberant environments. For simplicity, we assume that the localization is performed in the two-dimensional space where a position is described by a coordinate (x, y). The goal of this algorithm is to find the set of time indices for accurate position estimation. Previous studies used to regard the highest peak in $R^l(t)$ as the correct time lag for localization [7], [28] or introduced the peak quality to emphasize the highest peak [8], [9]. Focusing on detecting the preceding signal was an effective method to deal with multipath in the reverberant environments [20]. Although the above methods tried to achieve accurate localization, they treated each source independently. Since in ARL, the position is calculated based on all the source signals,

it is more desirable to determine the target position using all the multiple source signals jointly. For this, we divide the problem into the following two steps: minimizing the effect of multipath caused by reverberation and estimating the position via a reliability measure.

First, we attempt to minimize the multipath effects by carefully investigating the combinations of the time indices. Instead of picking only the highest peak as in (2.8), we need to select multiple peaks from $R^l(t)$. These peaks are selected from the local maxima of $|R^l(t)|$ that are greater than a predefined threshold γ. It is necessary to determine γ depending on the application because it makes a trade-off between computational complexity and accuracy since it controls the number of peaks to be selected. The time indices of selected peaks for the l-th source signal are denoted as $\mathcal{T}^l = \{\tau_p^l\}$ for $p = 1, \cdots, P^l$ where P^l denotes the number of peaks which varies depending on the characteristics of $R^l(t)$. If $R^l(t)$ has only a few prominent peaks, P^l will be small. On the contrary, if $R^l(t)$ is affected by severe reverberation or the target receiver is far from the source, P^l will have a large value. In this work, we set an upper bound ζ on the number of peaks P^l.

Next, we form a series of peak candidate set $\mathcal{P}_k$ which is a combination of time indices from each source with $k = 1, \cdots, K$. The total possible number of peak candidate sets will be $K = \prod_{l=1}^{L} P^l$ and the cardinality of $\mathcal{P}_k$ equals the number of sources $|\mathcal{P}_k| = L$. For instance, $\mathcal{P}_k$ can be described as $\mathcal{P}_k = \{\tau_a^1,\ \tau_b^2,\ \cdots, \tau_c^L\}$ where $\tau_a^1 \in \mathcal{T}^1, \tau_b^2 \in \mathcal{T}^2, \cdots, \tau_c^L \in \mathcal{T}^L$. An example of peak candidate set is depicted in the Fig. 4.1.

In each $\mathcal{P}_k$, we can select a combination of several time indices to derive the position of the target receiver. There are $n_p = \sum_{i=m}^{L} \binom{L}{i}$ possible combinations for each $\mathcal{P}_k$ where m is the minimum number of sources needed for the position

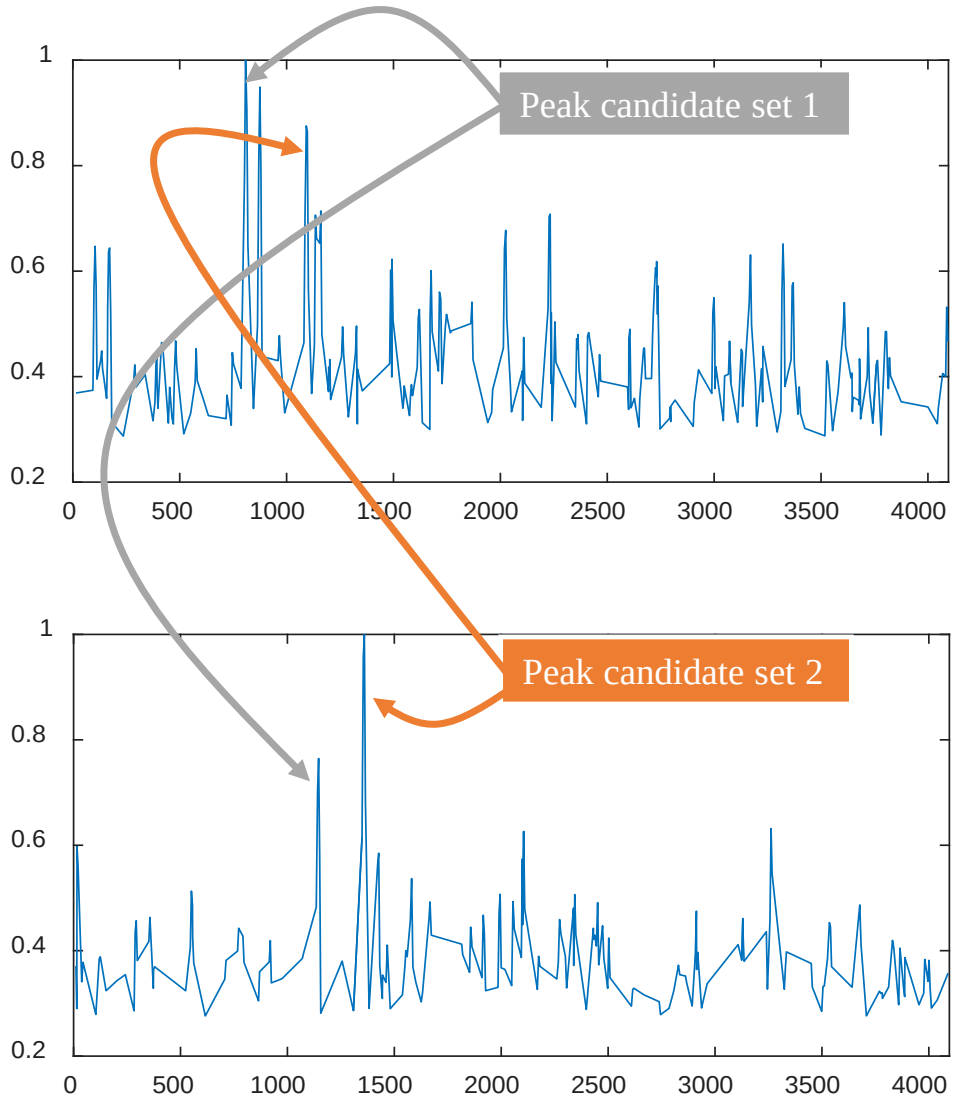


Figure 4.1. An example of peak candidate sets. A peak in cross-correlation of each source becomes a component in the peak candidate set.

estimation. For example, in a two-dimensional space, at least three time indices corresponding to three different sources are needed to calculate the position, i.e., $m = 3$. Let $(x_{k,n},\ y_{k,n})$ represent the position estimated from the n-th combination of $\mathcal{P}_k$ with $n = 1, \cdots, n_p$ using the LMA mentioned in the previous section. In order to measure how densely the estimated coordinates are populated, we calculate the variance of the n_p estimated positions as a degree of agreement in $\mathcal{P}_k$. We remove $\mathcal{P}_k$ from further consideration if

$$\sum_{n=1}^{n_p}(x_{k,n} - \bar{x}_k)^2 + \sum_{n=1}^{n_p}(y_{k,n} - \bar{y}_k)^2 > \nu \tag{4.1}$$

where $\bar{x}_k$ and $\bar{y}_k$ represent the mean of $\{x_{k,n}\}$ and $\{y_{k,n}\}$, respectively and ν is an empirically determined threshold. The relation between ν and the estimated position is depicted in the Fig. 4.2. Since the test (4.1) is applied to all $\mathcal{P}_k$, the L sound sources are jointly utilized to determine the target position.

This process enables to reduce the effect of early reflections which usually incur spurious peaks in $\{R^l(t)\}$. Also, we can benefit from weak sources because they contribute in finding valid candidate sets whereas previous studies often decided weak sources as unreliable and excluded them in estimating positions [8].

Even though the peak candidate sets are refined through the agreement test of (4.1), there still exists a possibility that some of the remaining candidate sets indicate irrelevant positions. In order to find these irrelevant candidate sets, we compute a reliability measure q_k for each $\mathcal{P}_k$. Some previous studies in source localization introduced various reliability measures that can be obtained from $R^l(t)$ [22], [51]. We have found that

$$q_k = \sum_{\tau_p^l \in \mathcal{P}_k} \frac{|R^l(\tau_p^l)|}{max(|R^l(t)|)} \tag{4.2}$$

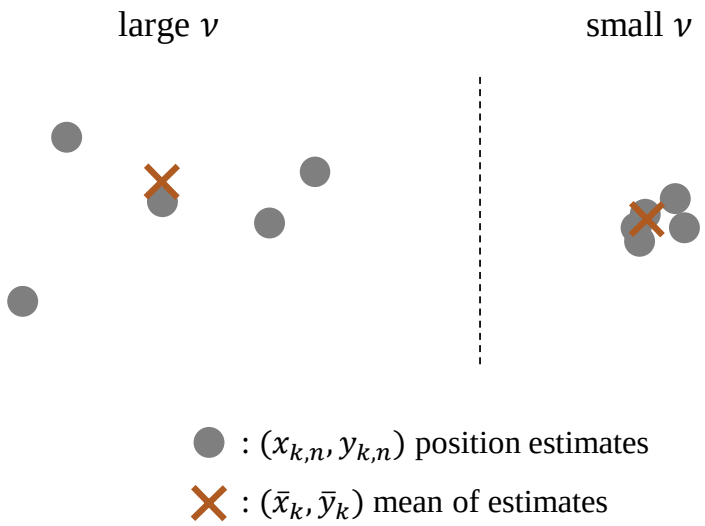


Figure 4.2. The relation of ν and the estimated position.

is suitable for measuring the reliability of $\mathcal{P}_k$. This reliability measure represents the confidence of each candidate's position while the previous studies focused on finding one salient peak from individual source. The final position $(\bar{x}_{\hat{k}},\ \bar{y}_{\hat{k}})$ is calculated from $\mathcal{P}_{\hat{k}}$ where

$$\hat{k} = \underset{k}{\operatorname{argmax}}\ q_k. \tag{4.3}$$

Since the reflective peaks or noise are likely to result in lower magnitude in $R^l(t)$, sorting $\mathcal{P}_k$ according to q_k will help to identify the most probable position among the candidate sets.

Table 4.1: System configuration

Parameter	**Value**
Sampling rate (kHz)	48
Frequency band (kHz)	18.0 - 20.9
Symbol length (samples)	8192

4.3 Performance evaluation

In this section, we compare the performance of three different TDE methods in ARL. The first method is the proposed robust TDE algorithm denoted by *RTDE*, which computes the receiver position using the multiple sources jointly. The second method denoted by *Prec* represents the technique introduced in [20], which considers the precedence of the highest peak if it exists. The last method denoted by *MaxPeak* applies the conventional highest peak detection method [7], [28] given in (2.8). All three methods were employed upon the same ARL system suggested in [20] with the same configuration as shown in Table 4.1.

In implementing the proposed TDE method, the peak selection threshold γ was empirically set to the half of the maximum magnitude of $R^l(t)$. The upper bound ζ of P^l was set to 20 and the threshold ν for the agreement test was set to 0.001.

The performances are depicted in terms of the localization error calculated by the root mean square error (RMSE) versus precision plots. The RMSE is calculated between the exact and estimated positions over the entire data symbols for each target position. In this work, the precision means the cumulative distribution of the localization error.

Table 4.2: Parameters of the real room environment

Parameter	**Value**
Dimension (m)	7.99 × 10.36 × 3.90
Loudspeaker position (m) (height: 1.14 m)	(0.87, 1.12), (6.92, 1.27), (6.57, 8.75), (1.69, 9.08)
RT_{60} (seconds)	0.571
Number of target positions	60

4.3.1 Performance evaluation in a real room

In order to evaluate the performance in reverberant environments, we conducted the experiments in a large reverberant seminar room in the presence of two experimenters and several rows of tables and chairs. The ambient noise measured in the environment was 35.3 dB. Four loudspeakers (Genelec 8030) were placed around the surrounding walls. The height of the target microphones (AKG C1000S) was similarly leveled with the loudspeakers at 1.20 m. The inaudible source signals were played simultaneously in a continuous manner from the loudspeakers. The average SPL of the source signals was measured as 46.4 dB at 1 m away from the loudspeakers. The reverberation time (RT_{60}) was measured using pink noise in a preliminary experiment [49]. The RT_{60} for considered part of frequency band was 0.335 seconds. The environmental parameters of this room are listed in Table 4.2. RMSEs were calculated at 60 different positions inside the room. The average of P^l for the real room experiment was 8.33.

From the results shown in Fig. 4.3, we can see that the proposed robust TDE algorithm outperformed the other methods in terms of both the precision and RMSE.

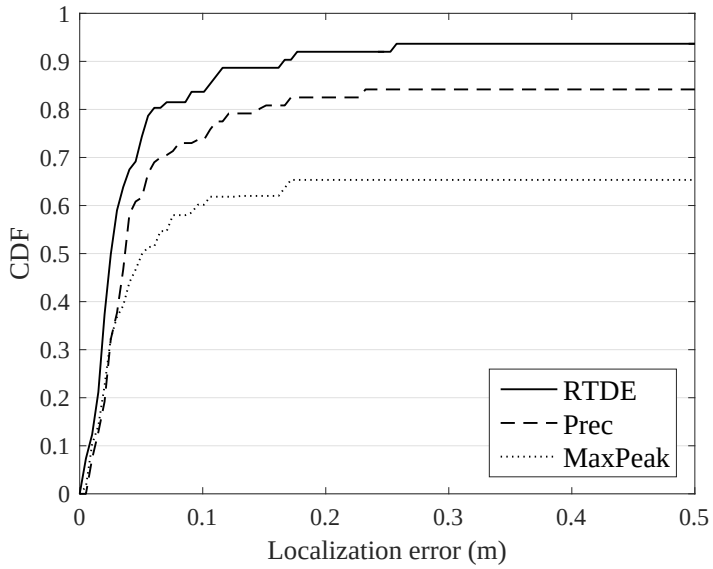
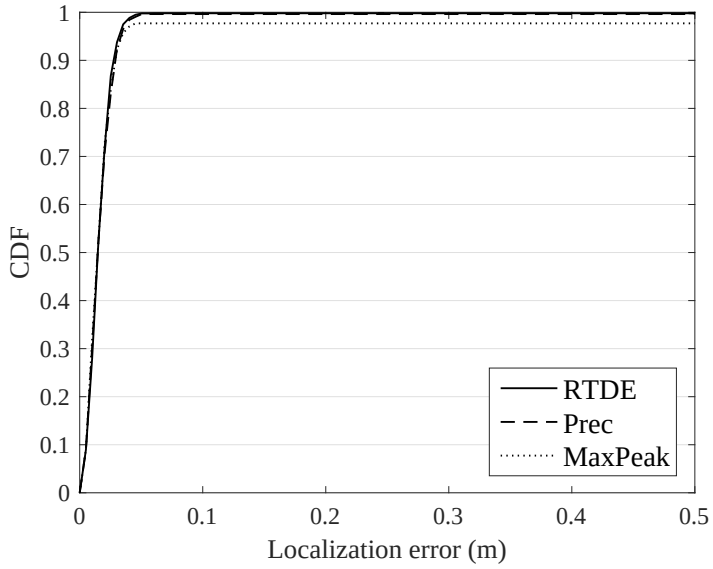


Figure 4.3. Performance evaluation of the ARL system using different TDE methods: comparison of the proposed method with the direct path detection and the conventional highest peak detection in the actual room environment.

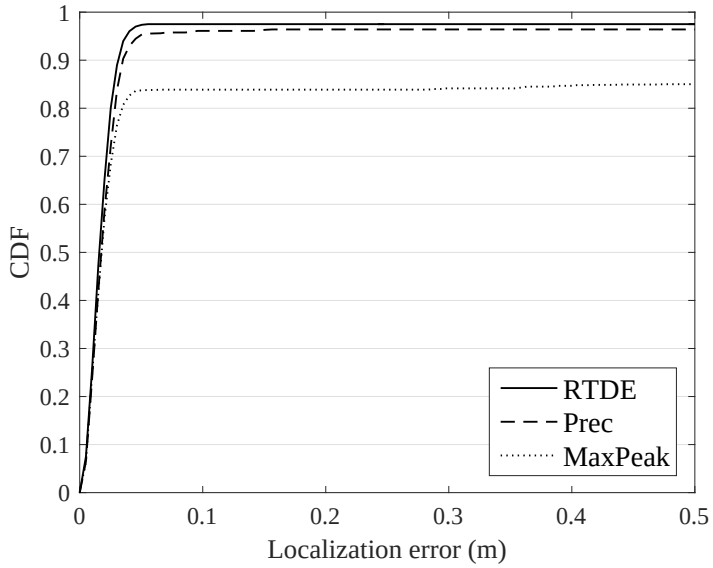
Showing better precision can be accounted for by the fact that analyzing the agreement of each peak candidate set jointly using the combinations of all sources is an effective way to disambiguate the detection process. With this comprehensive analysis, we can eliminate ambiguous or reflective peaks caused by multipath and select only the valid candidate sets. The reduction of RMSEs in the proposed method is considered to result from the sorting of the peak candidate sets according to the reliability measure.

4.3.2 Performance evaluation in simulated reverberant conditions

We also conducted experiments in simulated environments to evaluate the performance of the proposed method in various reverberant conditions. For this, we simu-

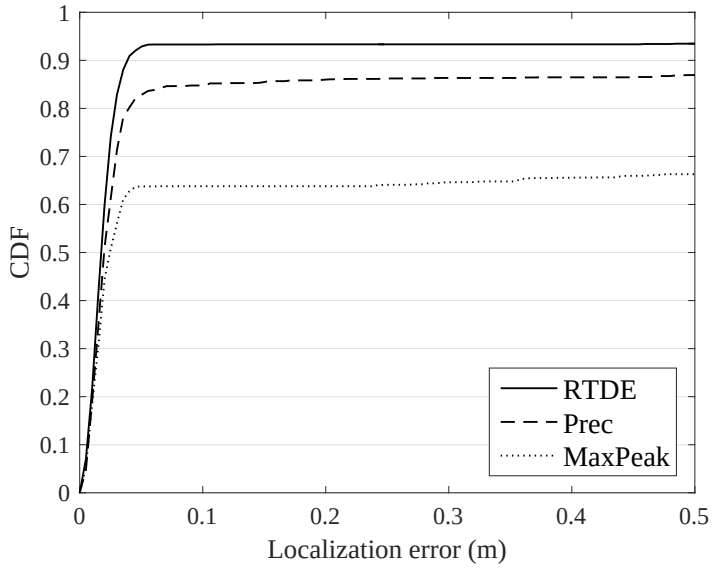


(a) RT = 0.3 seconds

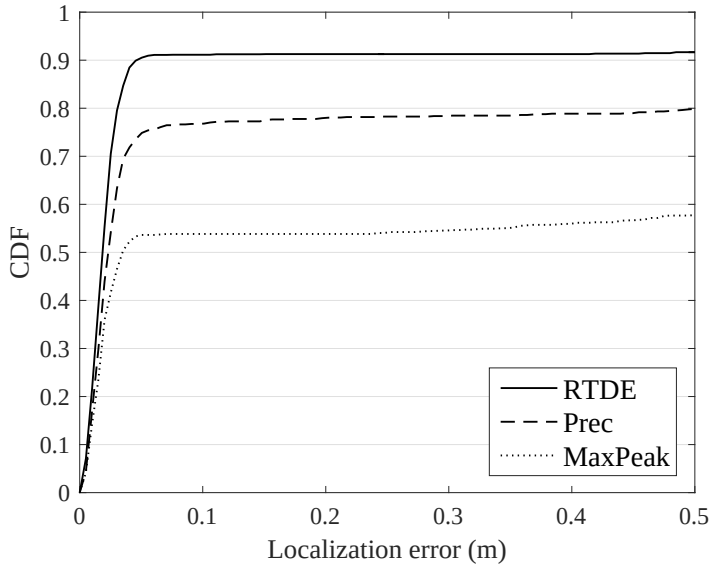


(b) RT = 0.5 seconds

Figure 4.4. Performance evaluation of the ARL system using different TDE methods in the simulated environments with the mild RT.



(a) RT = 0.8 seconds



(b) RT = 1.1 seconds

Figure 4.5. Performance evaluation of the ARL system using different TDE methods in the simulated environments with the high RT.

Table 4.3: Parameters of the simulated room environments

Parameter	**Value**
Dimension (m)	8×10
Loudspeaker position (m)	(1, 1), (7, 1), (1, 9), (7, 9)
RT_{60} (seconds)	0.3, 0.5, 0.8, 1.1
Number of target positions	3000

lated an enclosed space with the dimension and layout similar to those in the previous experiment while varying the RT_{60} as shown in Table 4.3. In order to simulate the environments, the RIRs corresponding to RT_{60} of 0.3, 0.5, 0.8 and 1.1 seconds were generated by the image method [48]. A total of 3000 target receiver positions were randomly chosen inside the simulated room and RMSEs were calculated. The configuration of the ARL system for this simulated environments was set the same to the previous real room experiment as shown in Table 4.1.

The results are shown in Figs. 4.4 and 4.5 where we can find that the proposed method outperformed the other approaches. It is worth noting that the proposed method showed rather consistent precision even the RT_{60} increased. This is because the proposed method evaluates a lot of possible combinations of time indices jointly whereas the conventional methods generally analyze time indices in a source-wise manner.

4.4 Summary

In this chapter, we have proposed a novel TDE algorithm for the ARL system that operates in real environments. It is difficult to estimate the correct time delays in

actual environments due to the multipath effect caused by acoustic reverberation. The proposed algorithm attempts to minimize the effect of multipath by investigating the multiple sources jointly and determines the location with the adoption of a reliability measure. The proposed algorithm showed better performance than the conventional methods in both the actual room and simulated environments with various reverberation conditions.

Chapter 5

Indoor Localization Based on Particle Filtering

5.1 Introduction

It is complicated to apply ARL in the actual environment since reverberation or reflections from various objects cause errors in the estimation of time delays. In highly reverberant environments, some reflections are often stronger than the direct path signal and the time delay estimation becomes more ambiguous [18], [19]. In the past, some researches focused on designing the data structure of the source for an application of indoor localization or estimating the accurate target position, rather than facing the reverberant environment. Previous studies conducted experiments in environments with low reverberations where space is surrounded by curtains [9] or a small designed space where direct path signals can be easily identified [7]. Recently, some researches began to deal with reverberation. Some studies [10], [20] tried to identify the direct path signal in reverberation. Studies in the SSL also fo-

cused on spurious measurements in the reverberant environments. DATEMM [21] tried to disambiguate TDOA measurements in the multiple source localization system and Canclini *et al.* [22] suggested a reliability criterion to disambiguate TDOA measurements in the reverberant condition.

In this chapter, we propose an indoor localization system for acoustic sensors based on the artificial sound sources through the particle filtering. Especially, we aimed to accurately estimate the position of the sensor with low-complexity in the environment with high reverberation. Reverberation is one of the biggest obstacles in position estimation using acoustic signals since it produces unrelated reflections. Although there are many acoustic localization systems, only a few studies aimed to deal with reverberation. We propose the acoustic receiver localization system with the peak quality, efficient likelihood function, and finding direct path peak region for accurate position estimation. Through these methods, we have improved the performance compared to the existing methods, and have verified the performance in various reverberation environments in both simulation and real environment.

5.2 A framework of positioning method using particle filter

This section describes the Bayesian framework of the positioning method using the particle filter. We define the dynamic model of the system, explain how to perform Bayesian estimation with the particle filter, and then introduce the formulation of the likelihood function. Since there are a number of famous articles regarding the particle filter, we address only the essentials in understanding our study. A full description of the particle filtering theory can be found in [52], [53].

5.2.1 State and dynamic models

In order to describe the system in the Bayesian framework, the dynamic model should be defined. There are several dynamic models for representing the position of people. Among them, the Langevin model which is proposed to represent the Brownian motion is widely used in practice because it represents the motion of people well even though its simple implementation [38], [39], [54]–[56]. In the Langevin model, the Cartesian coordinates are assumed to be independent to each other. We will consider it in two-dimensional space because the height of the loudspeakers and the receiver are similar in the system. Let us define the state of the target receiver at time t as $\mathbf{x}_t = [x_t \ y_t \ \dot{x}_t \ \dot{y}_t]$ where $[x_t \ y_t]$ represents the coordinate and $[\dot{x}_t \ \dot{y}_t]$ represents the velocity, then the Langevin motion model for x-coordinate can be written as:

$$x_t = x_{t-1} + \Delta T \dot{x}_{t-1}, \tag{5.1}$$

$$\dot{x}_t = a_x \dot{x}_{t-1} + b_x F_x \tag{5.2}$$

where $F_x = \mathcal{N}(0, 1)$ represents a normally distributed random variable with zero-mean and variance of one, and $\Delta T = L_f / f_s$ means the time interval with L_f being the frame length in samples and f_s denoting the sampling frequency. a_x and b_x are the position and velocity variance constants, respectively, which are defined as following:

$$a_x = e^{-\beta_x \Delta T}, \tag{5.3}$$

$$b_x = \dot{x}_t \sqrt{1 - {a_x}^2}. \tag{5.4}$$

Similar to previous studies [38], [39], [54]–[56], we set the rate constant $\beta_x = 10/s$ and the velocity parameter $\dot{x}_t = 1m/s$ which are found to show decent performance

for the indoor positioning.

5.2.2 Bayesian framework using particle filter

For the localization system with L loudspeakers, we will consider the peak indices of $R^l(t)$ in (2.7) as the measurements, which can be written as $Z_t^l = \{z_t^l\}$ where

$$z_t^l = \{z_{t,1}^l, \cdots, z_{t,P}^l\} \subseteq \{\tau_j^l\} \tag{5.5}$$

and P is the number of peaks from the result of the peak detection on $R^l(t)$. Let $Z_{1:t}$ be the measurements up to time t where the superscript l is omitted in this section. The state $\mathbf{x}_t$ will be assumed to follow the Markov process and independent to the measurements. The goal of the system is to estimate the posterior $p(\mathbf{x}_t|Z_{1:t})$, which can be calculated based on the Bayesian recursive estimation as following:

$$p(\mathbf{x}_t|Z_{1:t}) = \frac{p(Z_t|\mathbf{x}_t)p(\mathbf{x}_t|Z_{1:t-1})}{p(Z_t|Z_{1:t-1})} \tag{5.6}$$

$$p(Z_t|Z_{1:t-1}) = \int p(Z_t|\mathbf{x}_t)p(\mathbf{x}_t|Z_{1:t-1})d\mathbf{x}_t \tag{5.7}$$

where $p(\mathbf{x}_t|Z_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|Z_{1:t-1})d\mathbf{x}_{t-1}$ and $p(Z_t|\mathbf{x}_t)$ are the prior and the likelihood, respectively [55]–[57]. This implies that the probability distribution of the current state can be recursively obtained if there is a posterior of the previous time and the dynamic model. In general, there is no closed form solution, and we will estimate it through the particle filtering approach.

In the particle filter, the posterior at time t is expressed with particles $\{\mathbf{x}_t^{(i)}\}$ and their corresponding weights $\{w_t^{(i)}\}$ where $i = 1, \cdots, N$ is the index of the particles. We can divide the process of the particle filtering into two stages: the propagation stage and the update stage. In the propagation stage, the particles $\mathbf{x}_{t-1}^{(i)}$ of previous time are propagated into the current time based on the proposal distribution

$p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}, Z_t)$ which is described by the dynamic model. In the update stage, the weights of the particles are updated according to the likelihood from the measurements through the following equation:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \mathcal{L}(\mathbf{x}_t^{(i)}|Z_t). \tag{5.8}$$

The weights are then normalized to have summation of one by $\sum_i w_t^{(i)} = 1$. The resampling process is usually performed to prevent the degeneracy problem in which only some particles have meaningful weights. We used the low variance resampling algorithm for the resampling strategy [57].

5.2.3 Likelihood function

For a given state $\mathbf{x}_t$, the likelihood function computes the likelihood of the observed data. The likelihood function should be well formulated to reflect the characteristics of $R^l(t)$ to estimate the location. We use the similar function as the one used in previous studies [38]–[40], [54]. The likelihood function assumes that the observed data $z_{t,p}^l$ is the time delay of true position distorted by additive Gaussian noise. The likelihood function of l-th source is expressed as

$$\mathcal{L}(\mathbf{x}_t^{(i)}|Z_t^l) = \sum_{p=1}^{P} \alpha_{t,p}^l \mathcal{N}(z_{t,p}^l; \tau^l(\mathbf{x}_t^{(i)}), \sigma^2). \tag{5.9}$$

Here $\mathcal{N}(x; \mu, \sigma^2)$ means the Gaussian distribution with mean μ and variance σ^2 evaluated at x. The p-th measurements $z_{t,p}^l$ is the index from (5.5) and $\alpha_{t,p}^l$ is the corresponding normalized amplitude, respectively and $\tau^l(\mathbf{x}_t^{(i)})$ is the expected time delay of i-th particle which is computed by

$$\tau^l(\mathbf{x}_t^{(i)}) = \parallel \mathbf{s}^l - \mathbf{x}_t^{(i)} \parallel /c. \tag{5.10}$$

The likelihood is obtained by multiplying (5.9) of L sources by

$$\mathcal{L}(\mathbf{x}_t^{(i)}|Z_t) = \prod_{l=1}^{L} \mathcal{L}(\mathbf{x}_t^{(i)}|Z_t^l) \tag{5.11}$$

based on the assumption that measurements of individual pairs are independent.

However, since our system assumes asynchrony between the sources and the receiver, i.e. the emission time is unknown, we need to modify the likelihood function to incorporate the emission time. The following is the likelihood that can handle the unknown emission time:

$$\mathcal{L}(\mathbf{x}_t^{(i)}, \hat{\eta}_t^{m,(i)}|Z_t^l) = \sum_{p=1}^{P} \alpha_{t,p}^l \mathcal{N}(z_{t,p}^l; \tau^l(\mathbf{x}_t^{(i)}) + \hat{\eta}_t^{m,(i)}, \sigma^2), \tag{5.12}$$

$$\mathcal{L}(\mathbf{x}_t^{(i)}, \hat{\eta}_t^{m,(i)}|Z_t) = \prod_{l=1}^{L} \mathcal{L}(\mathbf{x}_t^{(i)}, \hat{\eta}_t^{m,(i)}|Z_t^l), \tag{5.13}$$

$$\mathcal{L}(\mathbf{x}_t^{(i)}|Z_t) = \underset{m}{max}[\mathcal{L}(\mathbf{x}_t^{(i)}, \hat{\eta}_t^{m,(i)}|Z_t)]. \tag{5.14}$$

The estimation of emission time $\hat{\eta}_t^{m,(i)}$ is calculated as following:

$$\hat{\eta}_t^{m,(i)} = z_{t,p_{max}}^m - \tau^m(\mathbf{x}_t^{(i)}) \tag{5.15}$$

where $m = 1, \cdots, L$ is the index of the sources and $z_{t,p_{max}}^m$ is the maximum peak index of m-th source. This equation assumes that the maximum peak value of at least one source is due to the direct path signal.

According to the modified likelihood function, if at least one $z_{t,p_{max}}^m$ is due to the direct path signal, the likelihood will have a high value. However, in a real environment with reverberation, wrong emission time estimates obtained from the spurious peaks may provide a higher likelihood. Therefore we need to force the likelihood to be computed away from the region with early or late reverberation. The proposed method in the next section will give the solution to this problem.

5.3 ARL in reverberant environment

In this section, we propose three novel techniques for ARL system based on the particle filters to deal with the high reverberation. First, we explain how to weight the peaks of $R^l(t)$, which will be called as peak quality hereafter. The peak quality is an indicator to determine the reliability of peaks in $R^l(t)$. Second, we modify the likelihood function for efficient and accurate computation. A lot of studies on the position estimation through the particle filter have been using the similar likelihood function as in (5.9) and (5.11). We will modify it to reduce the computational complexity while increasing the accuracy. The third is a process of finding the region that contains the direct path in $R^l(t)$. With this process, the likelihood is computed in the specified region to minimize the effect of reverberation and it also helps to estimate the emission time.

5.3.1 Peak quality

In the process of ARL, $R^l(t)$ is computed by the convolution of the source signal and the received signal and the shape of $R^l(t)$ becomes similar to the RIR. Therefore $R^l(t)$ will have some peaks due to the early reverberation and the late reverberation in addition to the peak by the direct path signal as in the RIR. Since the likelihood function is represented by the sum of the information of several peaks as in (5.12), it is necessary to give proper weights, for example, giving small weights to peaks due to reverberation and a larger weight to the direct path peak.

Here we redefine the normalized amplitude of each peak as the peak quality and

modify (5.12) to have a coefficient of λ as following:

$$\mathcal{L}(\mathbf{x}_t^{(i)}, \hat{\eta}_t^{m,(i)} | Z_t^l) = \sum_{p=1}^{P} (\alpha_{t,p}^l)^\lambda \mathcal{N}(z_{t,p}^l; \tau^l(\mathbf{x}_t^{(i)}) + \hat{\eta}_t^{m,(i)}, \sigma^2). \tag{5.16}$$

Although the amplitude is not the precise information for the direct path peak, still, the amplitudes of peaks near the direct path tends to be larger than those from the late reverberation or noise. In the previous study [40], (5.16) was calculated with $\lambda = 0$ without consideration of the amplitude of $R^l(t)$. In a recent study [20], [38], [39], [58], researchers put interests in peak amplitudes which are the case of $\lambda = 1$. From the preliminary experiment, however, we found that there is a limit to represent weights of the desired peaks only by the normalized amplitude which is $\lambda = 1$ as reverberation gets worse. This is because the amplitude of peaks due to noise and late reverberation get similar or the distinction between direct path and early reverberation becomes increasingly ambiguous. However, if large λ, such as infinity, is used, it will be similar to the case only using the maximum peak of $R^l(t)$. We confirmed that it is important to find the appropriate λ according to the environment. Experimental results will be shown in the following section.

5.3.2 Efficient calculation of the likelihood function

As can be seen on (5.16), the likelihood function is obtained as the weighted sum of Gaussian functions. As reverberation increases, a number of spurious peaks increases and so does their contribution to the likelihood. For this reason, we propose to calculate the likelihood only near the region adjacent to the delay computed from each particle: $\tau^l(\mathbf{x}_t^{(i)}) + \hat{\eta}_t^{m,(i)}$. The adjacent region can be obtained as following,

$$E_{t,p}^l = \{p| \ |\tau^l(\mathbf{x}_t^{(i)}) + \hat{\eta}_t^{m,(i)} - z_{t,p}^l| < D_a\} \tag{5.17}$$

The threshold of adjacency D_a is defined as average distance of the uniformly distributed particle in the given space, which can be represented as following,

$$D_a = \sqrt{uv/N} f_s/c \tag{5.18}$$

where u and v are the width and the length of the room if we assumed the space to be two-dimensional.

Then, we apply the region to the likelihood function as

$$\mathcal{L}(\mathbf{x}_t^{(i)}, \hat{\eta}_t^{m,(i)} | Z_t^l) = \sum_{p \in E_{t,p}^l} (\alpha_{t,p}^l)^\lambda \mathcal{N}(z_{t,p}^l; \tau^l(\mathbf{x}_t^{(i)}) + \hat{\eta}_t^{m,(i)}, \sigma^2). \tag{5.19}$$

By excluding measurements that are not related to each particle, computation of summation is reduced and more accurate likelihood can be obtained.

5.3.3 Finding the direct path region

We now introduce the method that excludes measurements that are assumed to be early and late reverberation and determines the region where the direct path signal belongs in $R^l(t)$. This method significantly reduces errors caused by reverberation. There are two possible cases of the peak location of direct path signal in the reverberant environment. First, as an ideal case, the peak with the maximum value in $R^l(t)$ is due to the direct path signal. This happens when the loudspeaker and the microphone are close together and are minimally affected by reverberation or reflections. The likelihood should be calculated in the region including the maximum of $R^l(t)$. The second case is when the direct path signal precedes the maximum of $R^l(t)$. In reverberant environments, early reverberation often has a larger peak than the direct path signal if the target microphone is near the walls or the direct path signal is blocked by obstacles [18], [19]. Previous studies [10], [20] tried to identify

Input:

Peak indices $\boldsymbol{z}_t^l = \{z_{t,p}^l\}$ from $R^l(t)$

Output:

Let B_t^l be the set of selected peak indices

$p_{max} = argmax_p(\alpha_{t,p}^l)^2$

$\hat{p} = p_{max}$

$B_t^l = \{\hat{p}\}$

while $(z_{t,\hat{p}}^l - z_{t,\hat{p}-1}^l) < D_c$ & $(z_{t,p_{max}}^l - z_{t,\hat{p}}^l) < D_{max}$ **do**

$B_t^l = B_t^l \bigcup \{\hat{p} - 1\}$

$\hat{p} = \hat{p} - 1$

end while

where $\alpha_{t,p}^l = R^l(z_{t,p}^l) / \sum_p R^l(z_{t,p}^l)$,

$D_c = 2df_s/c$ and

$D_{max} = \sqrt{u^2 + v^2} f_s/c$

return B_t^l

Figure 5.1. Process of finding the direct path region.

the preceding peak of the maximum of $R^l(t)$ as the direct path peak. However, identifying the preceding peak was not perfect in the previous studies due to the ambiguity the direct path peak decision.

We decided to find a region that includes direct path peak through the following algorithm in the Fig. 5.1. Here, the preceding peak of the maximum of $R^l(t)$ is iteratively analyzed by means of the maximum distance D_c of the first reflection and the direct path signal. This is done until there is no more preceding peak and while the maximum length of the set is less than D_{max}. Here, D_c is defined as twice of the distance between the source and the adjacent wall, and D_{max} is the maximum distance between the sources and the receiver that can have in the given space. In the end, the resulting region B_t^l will contain the index of the direct path signal. This method can be applied to any positioning algorithm that computes the cross-correlation between the source and the received signal which results the similar shape with the RIR.

The final likelihood function becomes as

$$\mathcal{L}(\mathbf{x}_t^{(i)}, \hat{\eta}_t^{m,(i)} | Z_t^l) = \sum_{p \in Q} (\alpha_{t,p}^l)^\lambda \mathcal{N}(z_{t,p}^l; \tau^l(\mathbf{x}_t^{(i)}) + \hat{\eta}_t^{m,(i)}, \sigma^2) \tag{5.20}$$

where the summation range is now modified to $Q = E_{t,p}^l \bigcap B_t^l$.

This method of selectively applying the measurements also helps to estimate the emission time. In the equation (5.13), the likelihood candidates are obtained using the estimated $\hat{\eta}_{t,(i)}^m$. If we use the whole measurement as in the other studies, even though the weight of particles can be controlled by the Gaussian function, wrong estimate of $\hat{\eta}_{t,(i)}^m$ can cause a large likelihood due to the early and late reverberation. Using the proposed method, it is possible to estimate the accurate likelihood by minimizing the effect of reverberation.

Table 5.1: Configuration of the source signals

Parameter	**Value**
Sampling rate (kHz)	48
Frequency band (kHz)	18.0 - 20.9
Frame length (samples)	8192

5.4 Performance evaluation

In this section, we conducted a series of experiments to evaluate the performance of the proposed system. It was investigated in various simulated reverberant environments. To simulate a large reverberant space, the RIRs are generated through the image method [48] with various reverberation times (RTs) and synthesized with the sources. The sources of the ARL system is based on [20] with the configuration shown in Table 5.1. The RT is the measure for reverberation representing the time required for the sound level to decrease by 60 dB. We conducted experiments in different parameters and compared the performance with the previous methods. After that, we compared the performance of the proposed method with previous methods in the actual environments.

Performance is expressed through the localization error between the actual and the estimated position for each target position. The localization error is measured at 10-th frame assuming the convergence of all the case using the particle filters. A sample example for the convergence of the localization error is shown on the Fig. 5.2 in condition of RT = 1.1 seconds and $\lambda = 4$ using 1000 particles. The localization errors at 5000 target positions are averaged in frame-wise for this example. The performances are depicted by the cumulative distribution function (CDF) of the

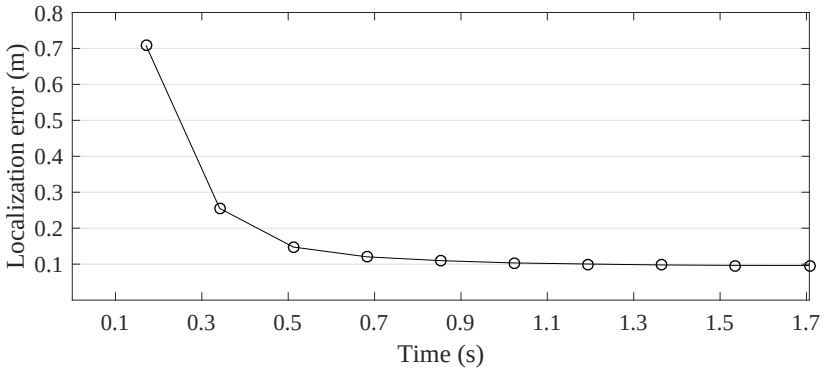


Figure 5.2. A sample experiment showing the convergence of the localization error in the case of RT = 1.1 seconds.

Table 5.2: Parameters of the simulated room environments

Parameter	**Value**
Dimension (m)	8×10
Loudspeaker position (m)	(1, 1), (7, 1), (1, 9), (7, 9)
RT (seconds)	0.5, 0.8, 1.1, 1.4
Number of target positions	5000

localization error and depicted only the region of interest.

5.4.1 Performance in a simulated environment

In order to measure the performance in various reverberant conditions, we conducted experiments in the simulated environments. The simulated RTs are 0.5, 0.8, 1.1, and 1.4 seconds. Experiments were performed on a total of 5000 arbitrary target positions. The σ^2 in the likelihood function was fixed to 10 for all experiments which showed good performance in the preliminary analysis. The configurations of

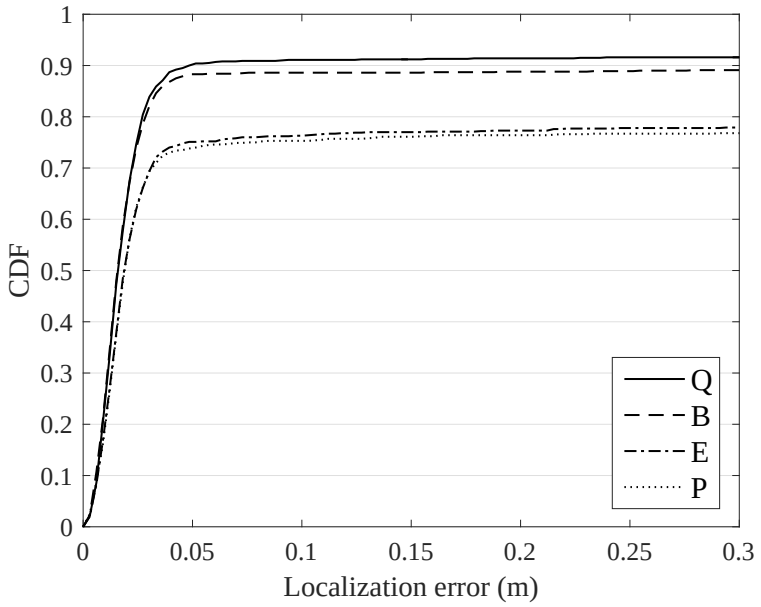


Figure 5.3. An experiment for the performance with/without the propose methods in RT = 1.1 seconds.

the simulated environment are listed on Table 5.2.

First of all, we introduce an experimental result to show the performance with and without the processes of efficient likelihood calculation and finding direct path region. The experimental result with the condition of RT = 1.1 seconds, 600 particles and $\lambda = 4$ is shown in the Fig. 5.3 to show the effect of the proposed techniques. Since we assume ARL system to be asynchronous, it is necessary to estimate the emission time. The conditions on Fig. 5.3 are as following,

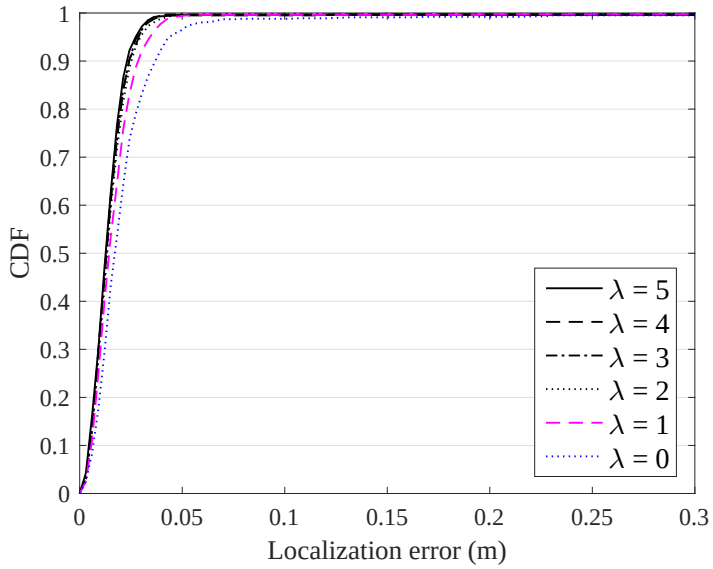
- P: the conventional method in (5.16),
- E: the application of an efficient likelihood as in (5.19),
- B: the application of finding the direct path region, and

- Q: the result when B and E applied together in (5.20).

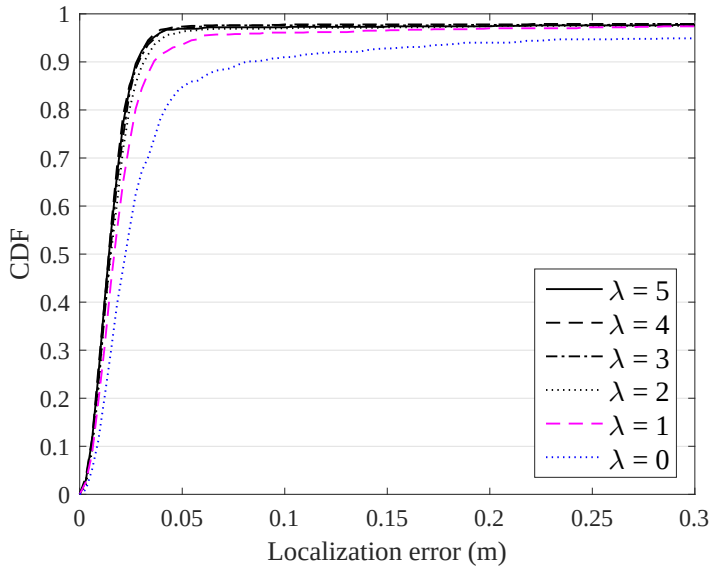
Although it may seem only slight improvement by the efficient likelihood calculation, the overall result shows evident performance improvement with the proposed methods. This is because the both techniques are necessary in estimating the emission time by limiting the contribution of spurious signals. Therefore, we applied Q through the following experiments.

We conducted experiments to investigate the effect of λ in the peak quality by changing the RT. The number of particles is fixed to 1000 and the result is shown in the Figs. 5.4 and 5.5. Here, $\lambda = 0$ is similar to the method used in [38], [40], and $\lambda = 1$ is similar to [39], [54], [58], [59]. From the results, it can be confirmed that the larger the value of λ, the higher the accuracy. This is because if the larger λ is used, it increase the weights on the reliable peaks, while the weight to the spurious peak decreases. However, performance does not continue to improve as λ increases. As λ increases, only peaks with high amplitude can contribute to the likelihood. If λ reaches the infinity, only the maximum peak will have the significant value and it will show the similar performance to the conventional method that estimates the location only through the maximum peaks as in (2.8). We plotted the failure rate in different λ on Fig. 5.6. In this plot, we regarded the target position with localization error more than 0.2 m as the failure and its ratio to the whole target positions are defined to be the failure rate. Therefore the value of λ needs to be decided according to the environment, here we proceeded with following experiments with $\lambda = 4$.

In the localization system, due to the nature of the application, there exists the degeneracy problem which means that only a few particles have the non-zero or meaningful weights. In compensation to the degeneracy problem, it is important to

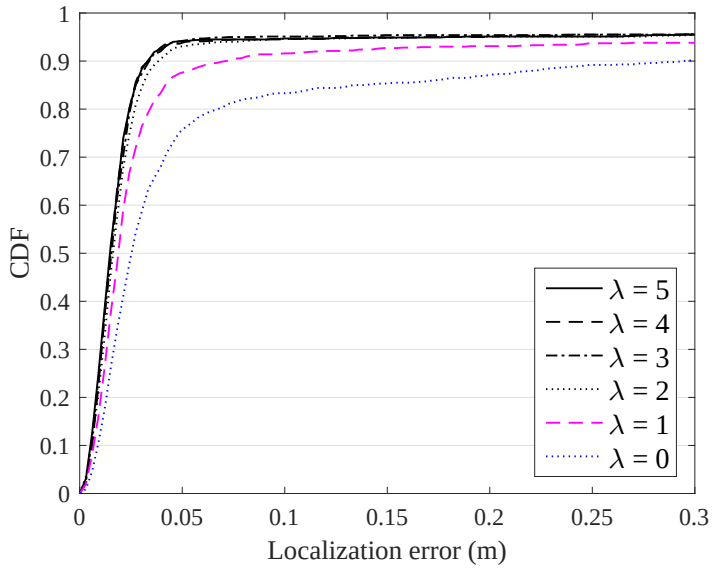


(a) RT = 0.5 seconds

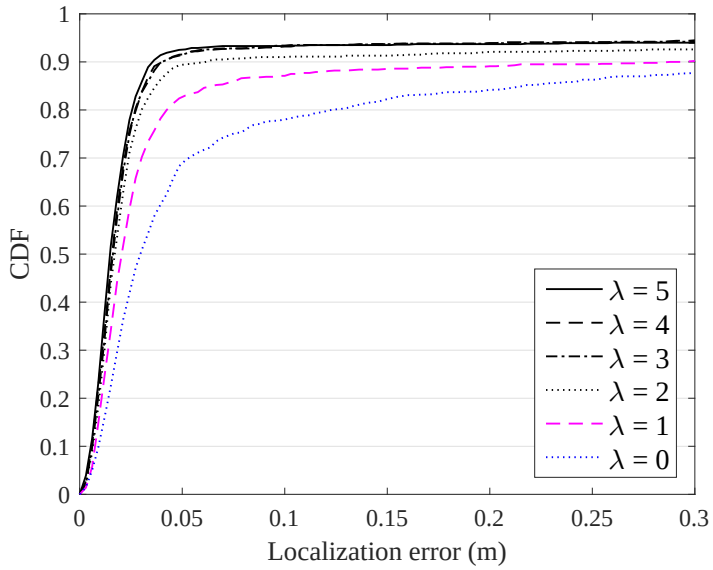


(b) RT = 0.8 seconds

Figure 5.4. Performance evaluation with regard to λ in the peak quality in mild RTs.



(a) RT = 1.1 seconds



(b) RT = 1.4 seconds

Figure 5.5. Performance evaluation with regard to λ in the peak quality in high RTs.

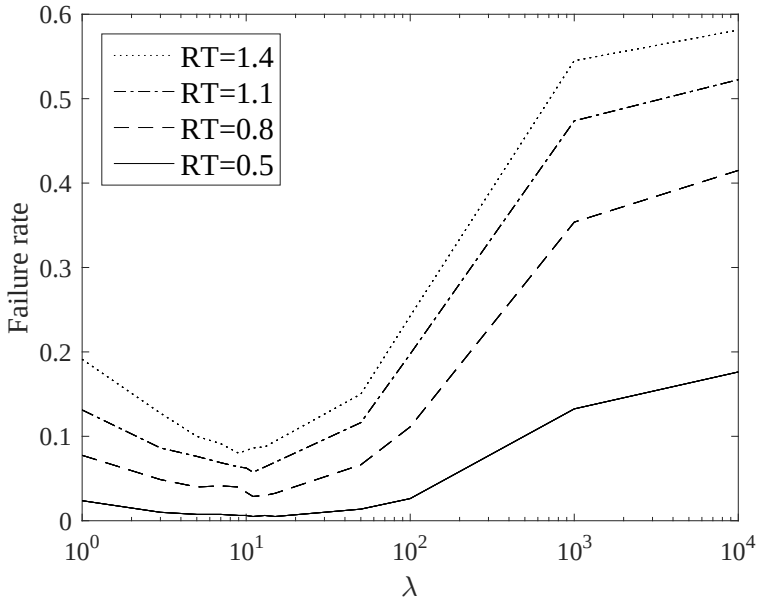
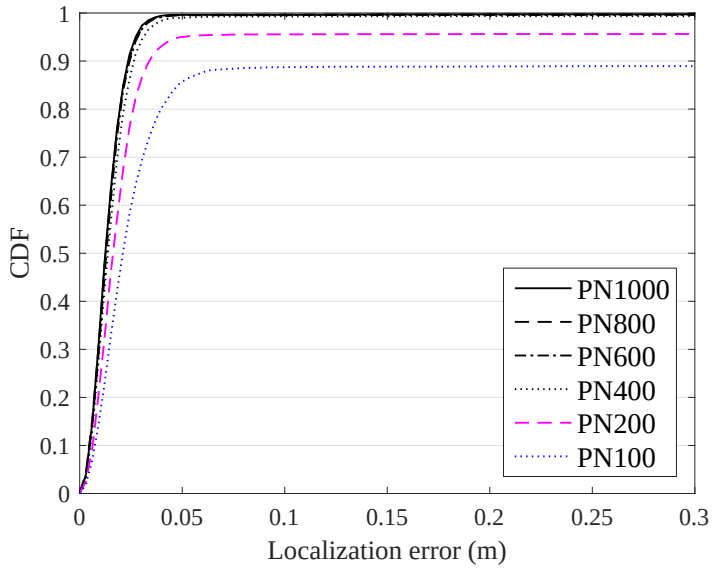


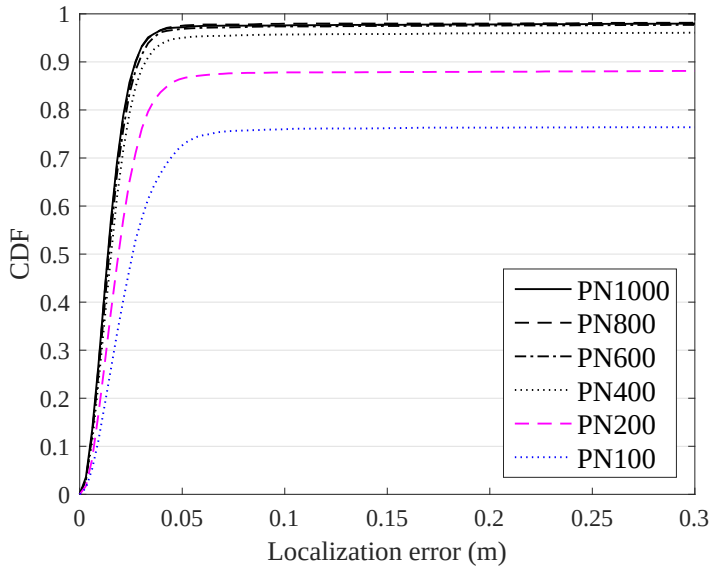
Figure 5.6. Failure rate of the proposed method with different λ.

choose a proper resampling strategy, but it is also important to check the number of particles suitable for the application. The particle filter shows accurate results as the number of particles increases, but there is a trade-off with the amount of computation. Therefore, we measured the performance by varying the number of particles (PN), which is shown in the Figs. 5.7 and 5.8. The figure shows that the performance converges around PN600.

Next, we conducted comparative experiments with other localization methods. Although there have been a number of studies, it is impractical to compare the performance with the same standard because the experimental environment and equipment are different. We compared performance with three other methods that can be conducted with the same standard. The first one is a conventional method of calculating the position based on TDOA through the LMA. This method works

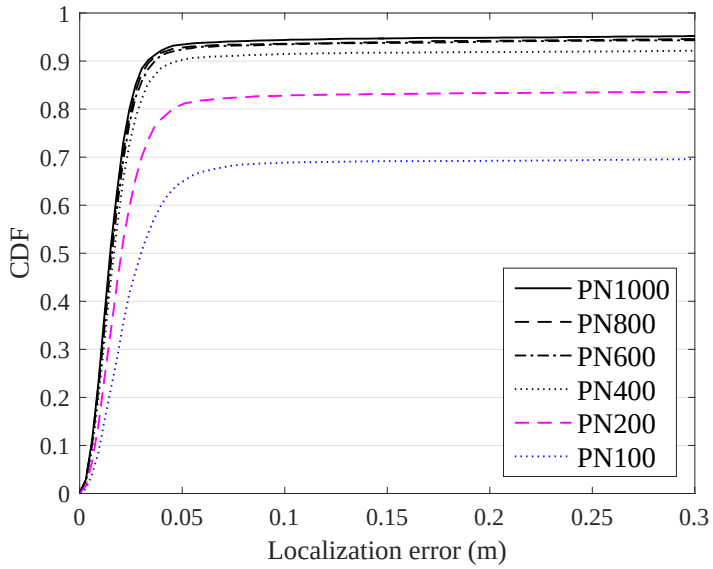


(a) RT = 0.5 seconds

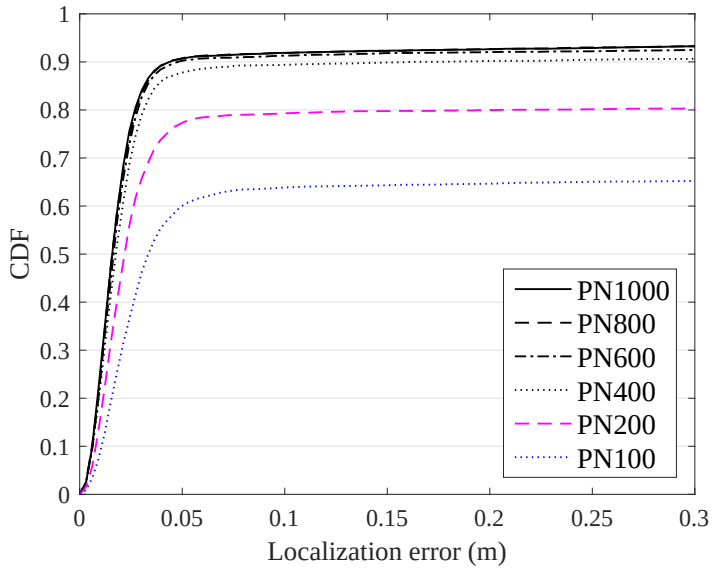


(b) RT = 0.8 seconds

Figure 5.7. Performance evaluation with regard to the particle number in mild RTs.



(a) RT = 1.1 seconds



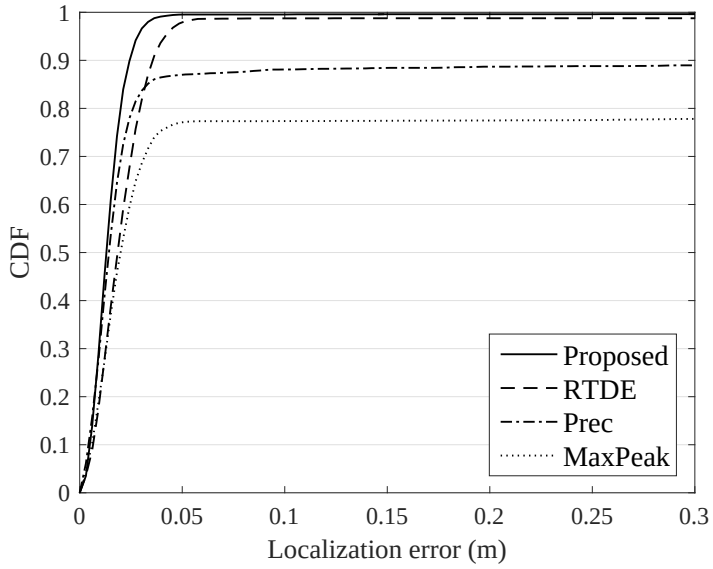
(b) RT = 1.4 seconds

Figure 5.8. Performance evaluation with regard to the particle number in high RTs.

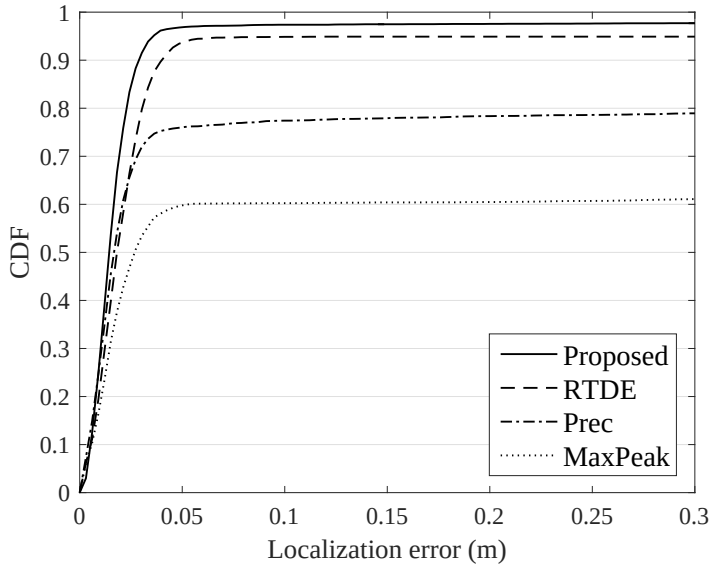
well in the ideal environment, but estimation will fail if multipath generates irrelevant peaks and induce wrong t_{prop}^{lm}. The second method is from [20], which finds the direct path peak among spurious peaks caused by multipaths. A similar scheme was introduced in the recent study [10]. However, this method also has a limitation that it does not work well in the high reverberation because finding the direct path peak gets more ambiguous. In severe reverberation, the multipath peaks are often larger than the direct path peak, resulting errors in TDE. The third method proposed in [59] aimed for accurate localization in severe reverberant environments. This method considered all possible t^{lm} candidates from combinations of peaks and estimates the position by ranking the peak candidates with a peak quality. Although its accurate estimation, too much computation is required as the number of candidates increases in high reverberations. We label the first, second and third method with *MaxPeak*, *Prec*, and *RTDE*, respectively. The performance comparison with the proposed method (*Proposed*) is shown in Figs. 5.9 and 5.10.

As the localization error measurements of the other methods, the RMSE of the average error of ten frames were calculated for each target position. We can see from Figs. 5.9 and 5.10, that the proposed method outperformed the other approaches. In the case of low reverberation (RT = 0.5 seconds), though the different accuracy, every method succeeds in locating more than over 75% targets. As the reverberation increases, the rate of failure increases rapidly for *MaxPeak* and *Prec*, and only *RTDE* and *Proposed* can estimate the position of most of the targets. The proposed method is not much different from *RTDE* in terms of CDF but the difference is more apparent as reverberation increases. Particularly, it is noteworthy that the proposed method has very little performance degradation due to reverberation.

Next, we compared the regional error of the proposed method with other con-

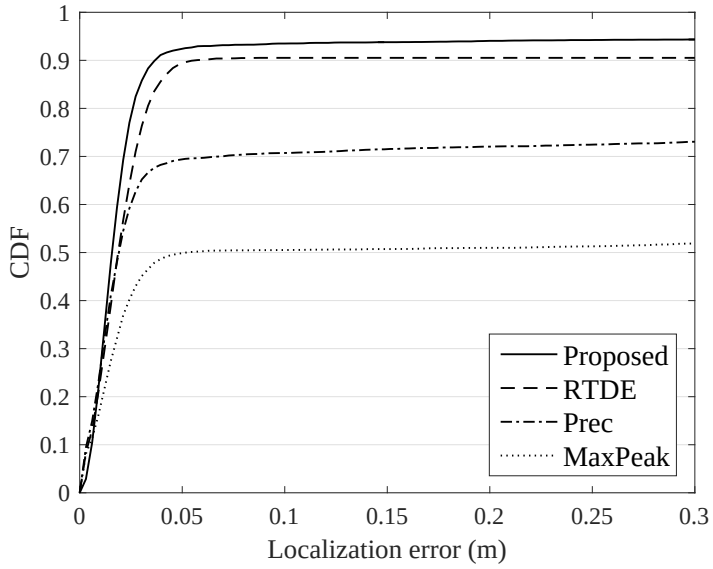


(a) RT = 0.5 seconds

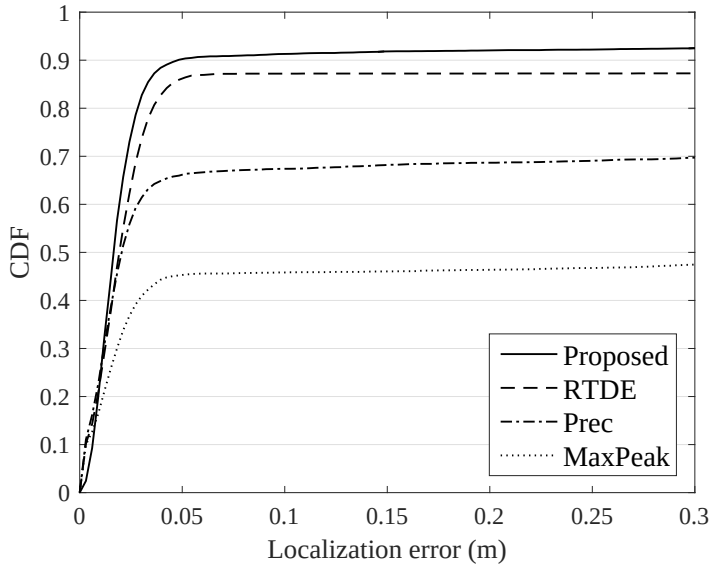


(b) RT = 0.8 seconds

Figure 5.9. Performance comparison of the proposed method with other methods in mild RTs.



(a) RT = 1.1 seconds



(b) RT = 1.4 seconds

Figure 5.10. Performance comparison of the proposed method with other methods in high RTs.

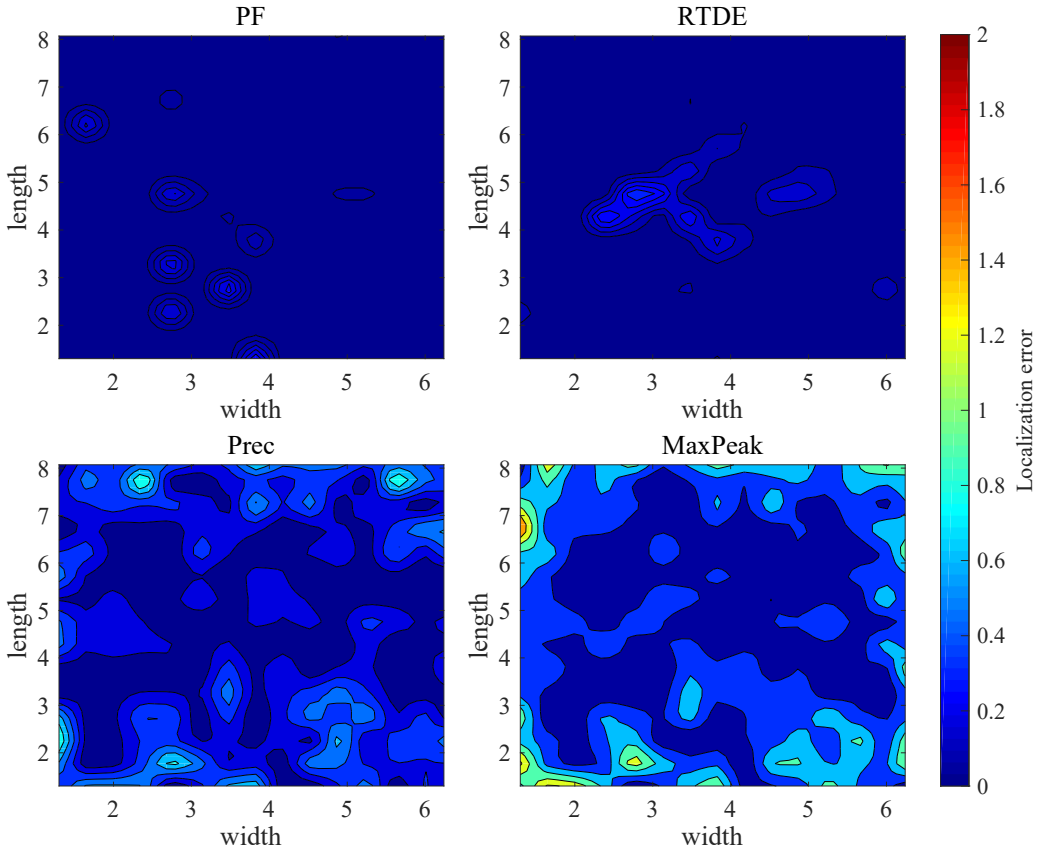


Figure 5.11. Regional error comparison of the proposed method with other methods when RT = 0.5.

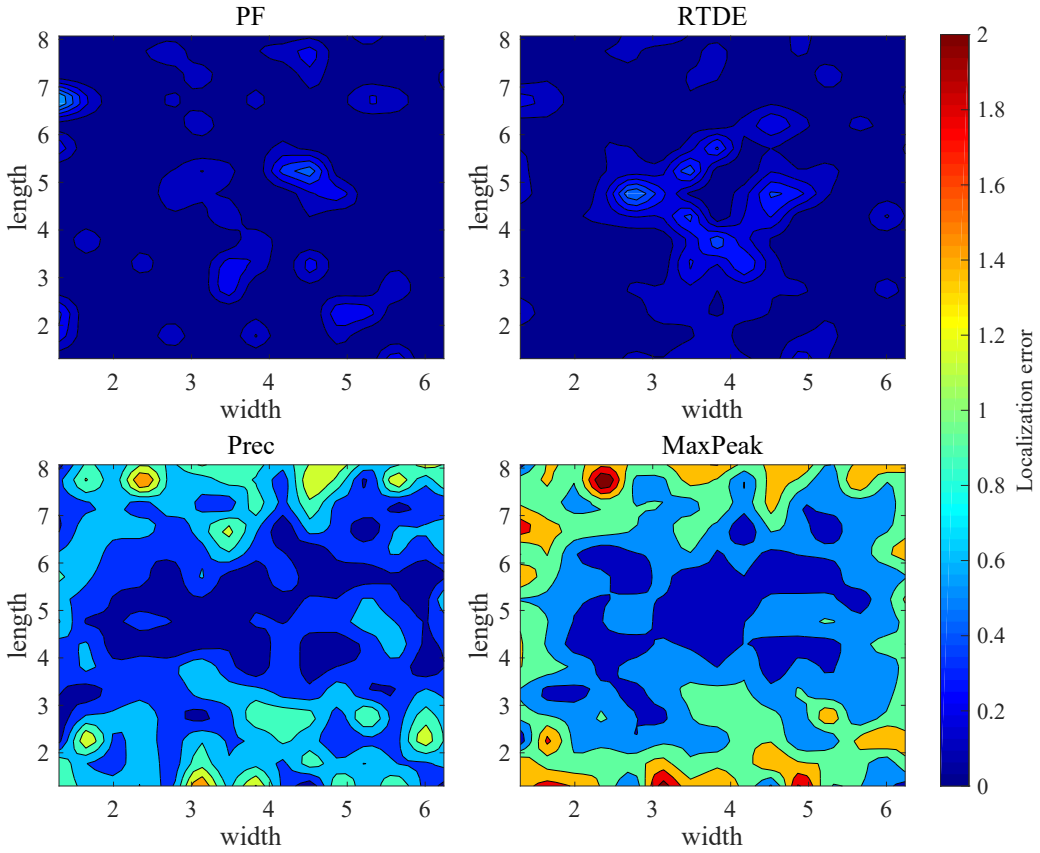


Figure 5.12. Regional error comparison of the proposed method with other methods when RT = 0.8.

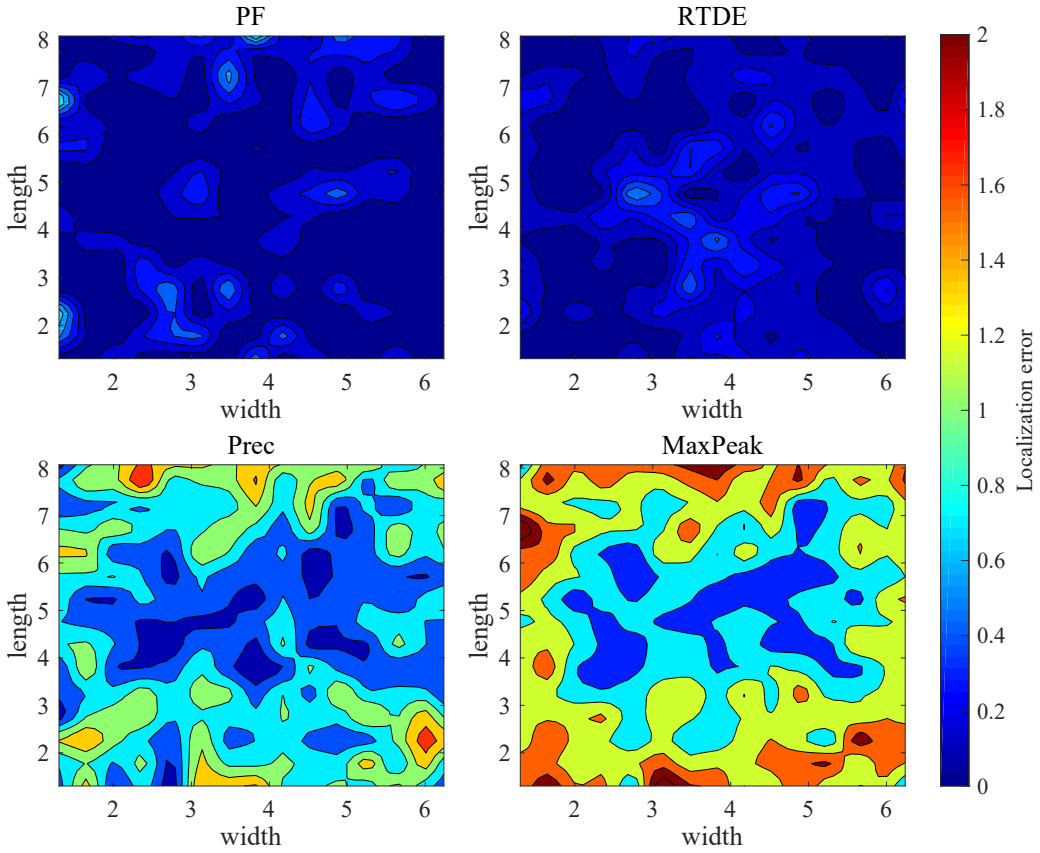


Figure 5.13. Regional error comparison of the proposed method with other methods when RT = 1.1.

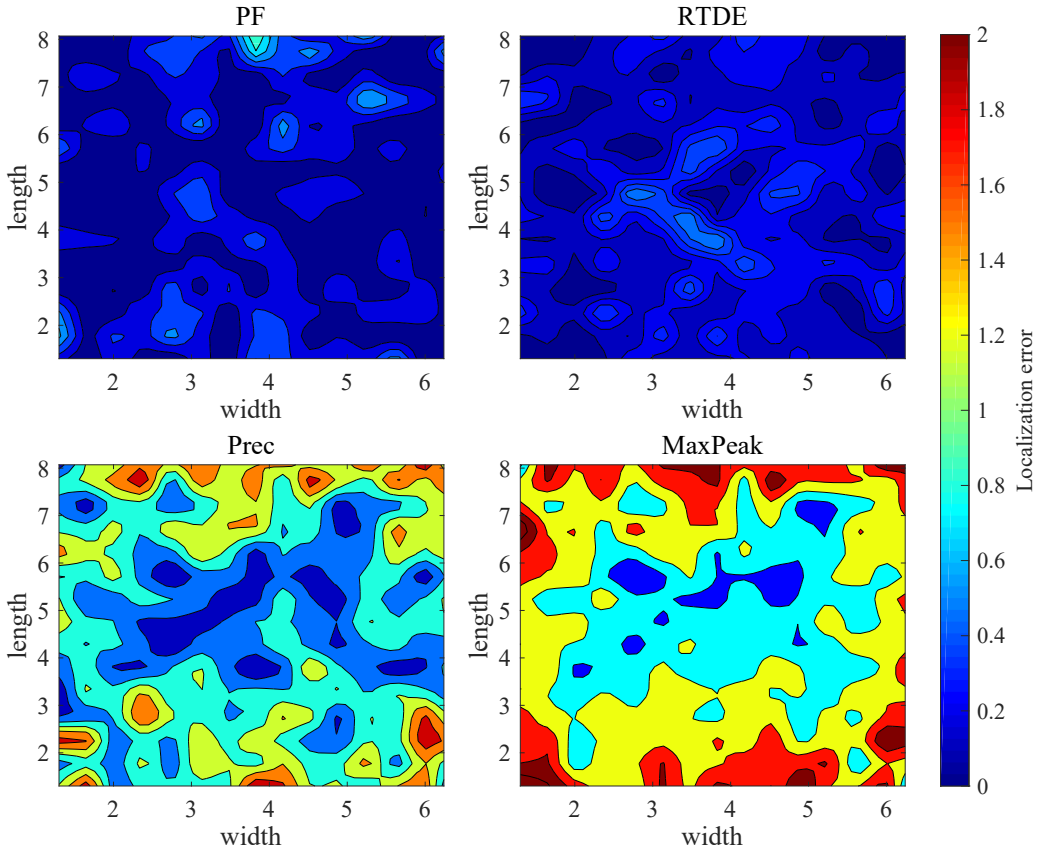


Figure 5.14. Regional error comparison of the proposed method with other methods when RT = 1.4.

ventional methods which is the contour of the equal localization error. Although we represented the performance with the cumulative distribution of localization error to see the performance in the given space, the localization errors can vary in inner area. We plotted the regional error in Figs. 5.11, 5.12, 5.13 and 5.14. In mild reverberation, all methods show similar performance over all area. However, in high reverberation conditions, we can see more localization error on the edge area in *MaxPeak* and *Prec* cases whereas the *Proposed* and *RTDE* shows rather consistent performance over all area. This is because the edge area is close to the walls which tend to have more reflections and conventional methods produce more errors when reflections are present.

We compared the calculation times for all the methods mentioned above. The computation time is expressed as the average of the time taken to estimate arbitrary 1000 target positions. Table 5.3 shows the relative computation time to the reference, i.e. the *MaxPeak* with time RT = 0.5 seconds. Computation time for the reference took 0.177 seconds using the Matlab program on the system with the i7-4770 processor and 8GB of RAM. In the case of *MaxPeak* and *Prec*, it takes almost the same time because the LMA is computed only once after TDOA is determined. Since *RTDE* requires the LMA and rank calculation for every peak candidate, it takes relatively longer time even with our best optimization. In contrast, the particle filter has a relatively small amount of computation. In particular, the PN600 showed a real-time calculation speed, showing possibility of practical use in the real applications.

Table 5.3: Comparison of computation time of the proposed method in various particle number with other methods.

RT	MaxPeak	Prec	RTDE	PN100	PN200	PN400	PN600	PN800	PN1000
0.5	1.000	1.005	39.233	1.676	2.903	4.366	5.910	7.413	9.012
0.8	1.121	1.127	45.069	1.694	2.702	4.486	6.135	7.778	9.717
1.1	1.397	1.406	45.249	1.742	2.807	5.527	6.408	7.746	9.798
1.4	1.289	1.300	45.564	1.764	3.080	4.660	6.374	7.957	10.104

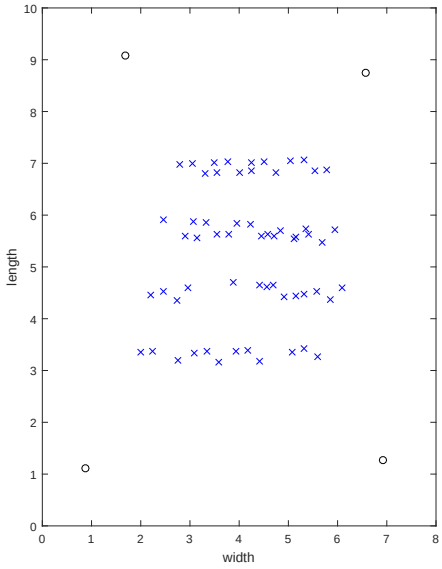
Table 5.4: Parameters of the two real room environments.

Parameter	**Classroom**	**Atrium**
Dimension (m)	7.99 × 10.36 × 3.90	20.95 × 15.83 × 2.68 (height up to 19.40 m at the center)
Loudspeaker position (m) (height: 1.14 m)	(0.87, 1.12), (6.92, 1.27), (6.57, 8.75), (1.69, 9.08)	(-6.29, -3.29), (-6.23, 3.33), (6.30, 3.23), (6.22, -3.27)
RT (seconds)	0.571	2.082
Ambient noise (dB)	35.3	39.1
Number of target positions	60	56

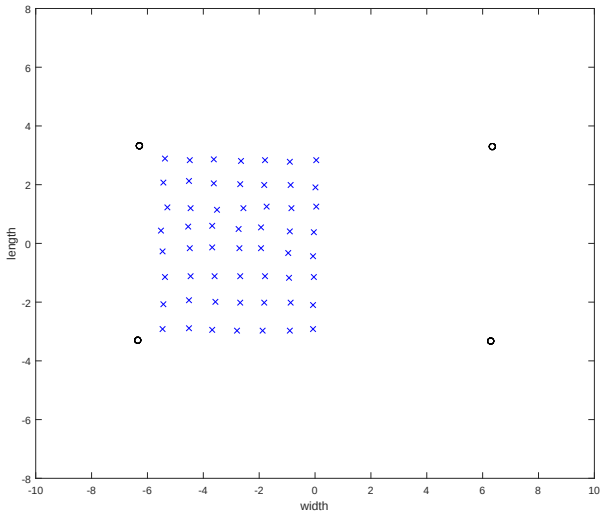
(a) Classroom

(b) Atrium

Figure 5.15. Two target places for experiments in the real environments.

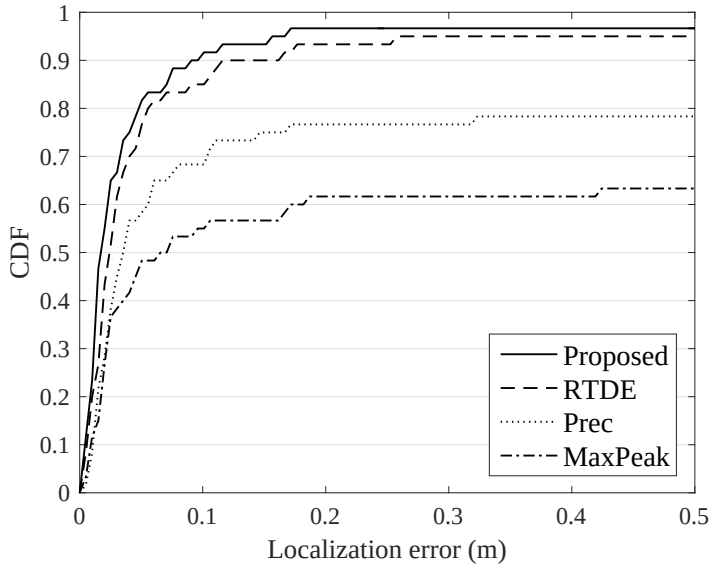


(a) Target positions on classroom

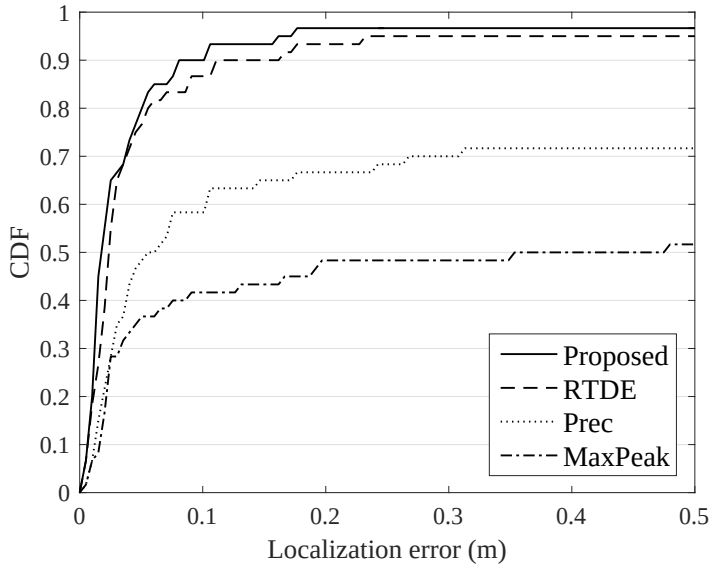


(b) Target positions on atrium

Figure 5.16. Target positions on the two given environments.

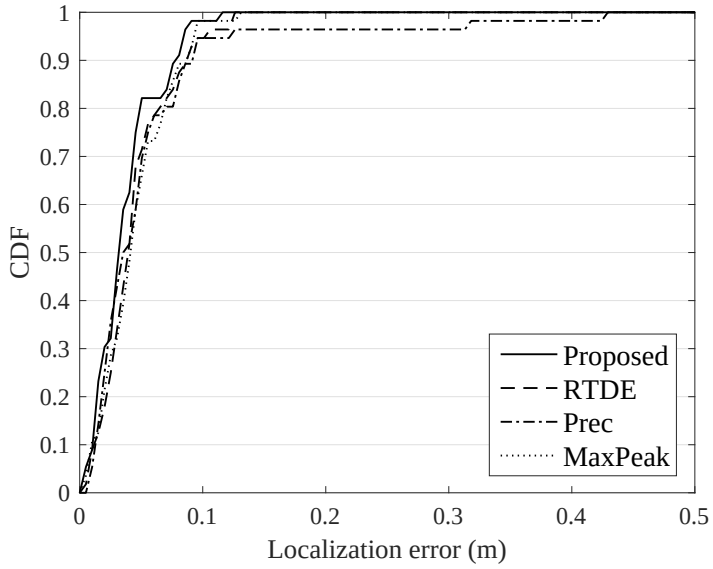


(a) LOS case

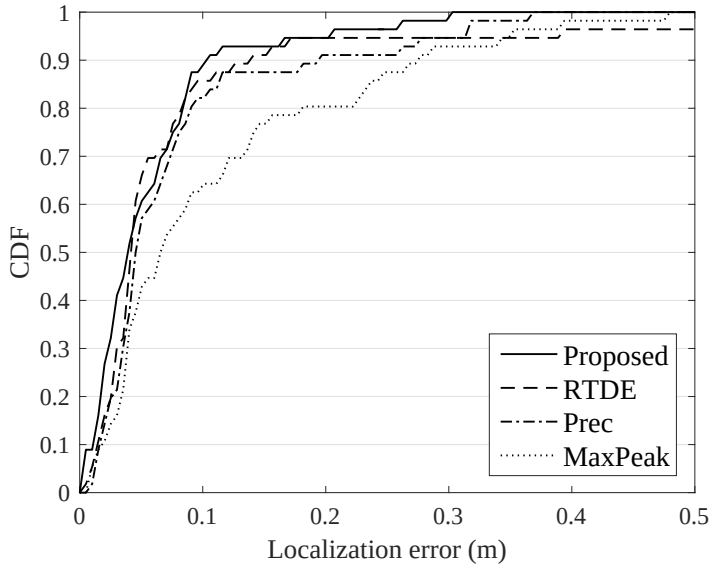


(b) NLOS case

Figure 5.17. Performance comparison of the proposed method and the other methods in the classroom environment.



(a) LOS case



(b) NLOS case

Figure 5.18. Performance comparison of the proposed method and the other methods in the main atrium environment.

5.4.2 Performance in the actual environment

We conducted experiments in two actual environments: a large and reverberant *classroom* and the wide-open *atrium*. For the large *classroom* case, experimental equipment for acoustic recording and playback, several rows of desks and chairs, and some furnitures were present in the room. Four loudspeakers (Genelec 8030) were installed near the wall around the room to reproduce the source signals. The microphone (AKG C1000S), which is the target of the position estimation, was placed inside the room in various positions with a height similar to loudspeakers. The performance is measured in total of 60 different target positions covering most of desk area as shown in Fig. 5.16a. The reference position of the microphone and the loudspeakers were measured by experimenters using the laser meter (Leica X310). The RT was measured to be 0.571 seconds in the preliminary experiment using the pink noise. The ambient noise at the site was measured to be 35.3 dB.

The *atrium* case is chosen since its space is larger and is more reverberant than the *classroom*. Four loudspeakers (Genelec 8030) were installed at each corner of the center area of the atrium. The center area of the atrium was about five stories high (19.4 m). The microphone (AKG C1000S) was positioned in a grid over various positions inside the room with a height similar to loudspeakers. By assuming symmetry, the performance was measured in total of 56 target positions (8 by 7 grid) only in half of the space. The positions of loudspeakers and target microphones are as shown in Fig. 5.16b. The RT was measured to be 2.082 seconds in the preliminary experiment using the sine sweep signal [49]. The ambient noise at the site was measured to be 39.1 dB. The configuration and parameters of both spaces are listed on Table 5.4. The λ is set to 4 and 1000 particles are used for the experiments.

We conducted experiments on two cases to investigate the performance in the environment with reverberation and reflections as well as the performance when the signal is blocked by an object. The first case refers to the LOS case which is the situation when there is no artificial object between the loudspeakers and the microphone considering only reverberation and reflection of room objects. The second case represents the non-line of sight (NLOS) that a person stands between the loudspeaker and the microphone with the shortest distance to block the dominant direct signal.

The performance of the LOS case and the NLOS case for two spaces are shown on Figs. 5.17 and 5.18, respectively. As we did in Figs. 5.9 and 5.10, we compared the performance in each situation with other methods. For *classroom* case, the proposed method outperformed other methods in the both cases as in the simulation. We can also see that the performance of the proposed algorithm and *RTDE* is consistent even in the *NLOS* situation where performance of other two methods(*MaxPeak* and *Prec*) degrades drastically. When the direct signal is blocked or attenuated, *RTDE* maintains its performance because it calculates for all peak combinations, but the other two methods have a high failure rate because they consider only one peak combination. The performance of the proposed method is better than the *RTDE* because it restricts the effect of reverberation as well as its low computation complexity.

For *atrium* case, the proposed method outperformed other methods in the both cases as well. Although the *atrium* is much more reverberant than the *classroom*, since the loudspeakers were placed away from the wall, the direct path signals were dominant. Therefore all methods in the LOS case showed better performance than the *classroom* case even with the severe reverberation. We can see the obvious per-

formance degradation of the *MaxPeak* case in the NLOS case.

From these results, we can deduce the possibility of using the proposed method in real life. In practice, since most built-in loudspeakers are installed on the ceiling, direct path signals will be dominant as in the *atrium* case. When the proposed method is used in the IoT environment, loudspeakers and microphones will be installed at various heights and locations. From the performance of NLOS cases in various environments, it can be considered to be applicable in the real environments.

5.5 Summary

In this chapter, we present an indoor acoustic receiver localization system that can operate in highly reverberant environments. To achieve low-complexity and high accuracy, we propose the system based on the particle filters. The likelihood function of the particle filter is modified to include the peak quality for accurate localization. In order to tackle the multipath effect, the proposed scheme employs a novel algorithm for finding the direct path region. The performance of the localization system is verified in a series of simulated reverberant environments, and also tested in the real environment. Experimental results show that the proposed method is superior to the conventional methods. However, due to the use of high-frequency inaudible sources, tracking has not been addressed in this paper due to the Doppler effect. We believe that the use of the particle filter will facilitate the extension to the receiver tracking as well, after further study on correlations in the presence of the Doppler effect.

Chapter 6

Conclusions

In this thesis, we designed an indoor positioning system using acoustic signals that operate in the real environments. The goal of the acoustic indoor positioning system is to find the position of a target acoustic sensor (i.e. a microphone) by reproducing a unique signal from loudspeakers with known positions. In addition, this system has advantages of being able to use the pre-installed loudspeakers and microphones on mobile devices. We have considered practical problems such as reverberation and the near-far effect for operation in the real environment. A series of proposed methods have been proposed and evaluated through experiments in the simulated and the real environments.

First, we designed the source data structure for a practical use. The proposed structure aims to work in the large reverberant space. In order to affect the least to the human hearing, the signals are carried through inaudible frequency bands. We borrowed a multiple access scheme called OFDMA-CDM in the wireless communication and modified it for the localization purpose. The proposed structure outperformed CDMA-based structure because it avoids the near-far effects. In the real

environment, due to the reflection and attenuation, reverberation is often stronger than the direct propagation signal. Since failure in detecting the direct propagation signal leads to wrong position estimation, we proposed the method of finding the direct path signal. The performance comparison between the traditional method showed that the proposed method outperformed in the reverberant environments.

Second, we proposed an algorithm for accurate localization in the highly reverberant environments. As the reverberation gets severe, the more spurious peaks appear in the RIR. The early reverberation makes it difficult the find the direct path signal and the late reverberation works as noise so that the time delay estimation gets ambiguous. The proposed method picks time delays from each source which are the indices of peaks in the cross-correlation and forms the peak candidate sets. The agreement test filters out inaccurate peak candidates set. Then the reliability is calculated in each peak candidate set and they are ranked by their reliability measure. The position is estimated from the most reliable peak candidate set. Unlike the previous localization method that estimates the position by one time delay estimate from each source, the proposed method takes into account all the possible combination which makes it more accurate in the reverberation condition and the case of the attenuated direct path signal. The performance of the proposed method is evaluated in the simulated and the real environments.

Third, we proposed to accurately estimate the position with relatively low computational complexity. Although the previous method provides the accurate estimation in the severe reverberation, it has disadvantages in high computational complexity. We propose to estimate the position with the particle filters that estimates the position by weighted particles. The weights of particles are computed by the likelihood. We proposed to form the likelihood function to reflect the reliability of

time delay observation and to be efficient. In addition, we proposed the process of finding the region that includes the direct path signal and computes likelihood in it. This process allows not only the accurate positioning but also estimation of unknown emission time in the asynchronous system. The performance of the proposed particle filter-based localization method is evaluated in the simulated environment by altering various parameters. Also, the performance in the real environments is evaluated for the line-of-sight and the non-line-of-sight case.

We proposed techniques that can be used in real life through a series of proposed methods and experiments. For example, it is possible to estimate the position of a smart device by using built-in loudspeakers or to estimate the position of home appliances in a home environment with multiple IoT devices. This technology can be used commercially, for example, by advertising, etc., or can be used for initial position calibration for accurate indoor LBS. Also, it can help to localize the target in the environment without LOS or non-visual situation where other sensors such as camera can't be used.

Although the proposed techniques are able to estimate the indoor position accurately and practically, there is a limitation in tracking the receiver. This can be summarized through the Doppler effect: when the receiver moves, the frequency shift occurs. Since the system uses narrow band high frequency that is shared by several sources, it is greatly influenced by the frequency shift even in the speed of human walking. Future study to tackle moving receivers can make this technique more suitable for practical use.

Bibliography

[1] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, Nov 2007.

[2] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, p. 707, 2016.

[3] Y. Huang, J. Benesty, and G. W. Elko, *Source Localization*. Boston, MA: Springer US, 2004, pp. 229–253.

[4] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO signal processing*. Springer Science & Business Media, 2006.

[5] I. Cohen, J. Benesty, and S. Gannot, *Speech processing in modern communication: Challenges and perspectives*. Springer Science & Business Media, 2009, vol. 3.

[6] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*. Springer Science & Business Media, 2008, vol. 1.

[7] N. Aloui, K. Raoof, A. Bouallegue, S. Letourneur, and S. Zaibi, "Performance evaluation of an acoustic indoor localization system based on a fingerprinting technique," *EURASIP J. Advances Signal Process.*, vol. 2014, no. 1, pp. 1–16, Jan. 2014.

[8] I. Rishabh, D. Kimber, and J. Adcock, "Indoor localization using controlled ambient sounds," in *Proc. Int. Indoor Positioning and Indoor Navigation (IPIN) Conf.*, 2012, pp. 1–10.

[9] C. Sertatıl, M. A. Altınkaya, and K. Raoof, "A novel acoustic indoor localization system employing CDMA," *Digital Signal Process.*, vol. 22, no. 3, pp. 506–517, May 2012.

[10] D. B. Haddad, W. A. Martins, M. d. V. M. da Costa, L. W. P. Biscainho, L. O. Nunes, and B. Lee, "Robust acoustic self-localization of mobile devices," *IEEE Transactions on Mobile Computing*, vol. 15, no. 4, pp. 982–995, April 2016.

[11] J. Li, G. Han, C. Zhu, and G. Sun, "An indoor ultrasonic positioning system based on toa for internet of things," *Mobile Information Systems*, vol. 2016, 2016.

[12] F. Seco, J. C. Prieto, A. R. J. Ruiz, and J. Guevara, "Compensation of multiple access interference effects in cdma-based acoustic positioning systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 10, pp. 2368–2378, Oct 2014.

[13] A. Plinge, F. Jacob, R. Haeb-Umbach, and G. A. Fink, "Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-

art algorithms," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 14–29, July 2016.

[14] S. Thrun, "Affine structure from sound," in *NIPS*, 2005, pp. 1353–1360.

[15] L. Wang, T.-K. Hon, J. D. Reiss, and A. Cavallaro, "Self-localization of ad-hoc arrays using time difference of arrivals," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 1018–1033, 2016.

[16] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens, "Auto-localization in ad-hoc microphone arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on.* IEEE, 2013, pp. 106–110.

[17] M. Cobos, J. J. Perez-Solano, Ó. Belmonte, G. Ramos, and A. M. Torres, "Simultaneous ranging and self-positioning in unsynchronized wireless acoustic sensor networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 22, pp. 5993–6004, 2016.

[18] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *J. Acoust. Soc. Amer.*, vol. 106, no. 4, pp. 1633–1654, Oct. 1999.

[19] K. W. Wilson and T. Darrell, "Learning a precedence effect-like weighting function for the generalized cross-correlation framework," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 2156–2164, Nov. 2006.

[20] J. Choi, J. Kim, S. J. Kang, and N. S. Kim, "Reverberation-robust acoustic indoor localization," in *Proc. Interspeech*, Sept. 2015, pp. 3422–3425.

[21] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 8, pp. 1479–1489, Nov. 2008.

[22] A. Canclini, P. Bestagini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro, "A robust and low-complexity source localization algorithm for asynchronous distributed microphone networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 10, pp. 1563–1575, Oct. 2015.

[23] K. Fazel and S. Kaiser, *Multi-Carrier and Spread Spectrum Systems: From OFDM and MC-CDMA to LTE and WiMAX.* Wiley, 2008.

[24] W. Y. Y. C. G. K. Yong Soo Cho, Jaekwon Kim, *MIMO-OFDM Wireless Communications with MATLAB.* Wiley, 2010.

[25] H. Schulze and C. Lueders, *Theory and Applications of OFDM and CDMA: Wideband Wireless Communications.* Wiley, 2005.

[26] J. Proakis and M. Salehi, *Digital Communications, 5th Edition*, 5th ed. McGraw-Hill, Nov. 2007.

[27] J. C. Prieto, A. R. Jimenez, J. Guevara, J. L. Ealo, F. Seco, J. O. Roa, and F. Ramos, "Performance evaluation of 3d-locus advanced acoustic lps," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 8, pp. 2385–2395, Aug 2009.

[28] F. Hoflinger, R. Zhang, J. Hoppe, A. Bannoura, L. Reindl, J. Wendeberg, M. Buhrer, and C. Schindelhauer, "Acoustic self-calibrating system for indoor smartphone tracking (ASSIST)," in *Proc. Int. Indoor Positioning and Indoor Navigation (IPIN) Conf.*, 2012, pp. 1–9.

[29] P. Lazik and A. Rowe, "Indoor pseudo-ranging of mobile devices using ultrasonic chirps," in *Proc. 10th ACM Conf. Embedded Network Sensor Syst.*, 2012, pp. 99–112.

[30] A. Ens, F. Höflinger, J. Wendeberg, J. Hoppe, R. Zhang, A. Bannoura, L. M. Reindl, and C. Schindelhauer, "Acoustic self-calibrating system for indoor smart phone tracking," *International Journal of Navigation and Observation*, vol. 2015, 2015.

[31] H. Lee, T. H. Kim, J. W. Choi, and S. Choi, "Chirp signal-based aerial acoustic communication for smart devices," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, April 2015, pp. 2407–2415.

[32] J. Wendeberg, J. Müller, C. Schindelhauer, and W. Burgard, "Robust tracking of a mobile beacon using time differences of arrival with simultaneous calibration of receiver positions," in *Proc. Int. Indoor Positioning and Indoor Navigation (IPIN) Conf.*, Nov 2012, pp. 1–10.

[33] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.

[34] G. Shen, R. Zetik, and R. Thoma, "Performance comparison of TOA and TDOA based location estimation algorithms in LOS environment," in *Proc. Workshop on Positioning, Navigation and Commu. (WPNC)*, Mar. 2008, pp. 71–78.

[35] C. Mensing and S. Plass, "Positioning algorithms for cellular networks using TDOA," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 4, May 2006, pp. 513–516.

[36] K. Yu, J. P. Montillet, A. Rabbachin, P. Cheong, and I. Oppermann, "UWB location and tracking for wireless embedded networks," *Signal Process.*, vol. 86, no. 9, pp. 2153–2171, Sept. 2006.

[37] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis*. Springer, 1978, pp. 105–116.

[38] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 826–836, Nov 2003.

[39] X. Zhong and J. R. Hopgood, "Particle filtering for tdoa based acoustic source tracking: Nonconcurrent multiple talkers," *Signal processing*, vol. 96, pp. 382–394, 2014.

[40] J. Vermaak and A. Blake, "Nonlinear filtering for speaker tracking in noisy and reverberant environments," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, vol. 5, 2001, pp. 3021–3024.

[41] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer Berlin Heidelberg, 2007.

[42] H. Kuttruff, *Room acoustics*. Crc Press, 2016.

[43] H. Kuttruff, *Acoustics: an introduction*. CRC Press, 2007.

[44] T. Aguilera, F. J. Álvarez, A. Sánchez, D. F. Albuquerque, J. M. N. Vieira, and S. I. Lopes, "Characterization of the near-far problem in a cdma-based acous-

tic localization system," in *2015 IEEE International Conference on Industrial Technology (ICIT)*, March 2015, pp. 3404–3411.

[45] S. Kaiser, "OFDM code-division multiplexing in fading channels," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1266–1273, Aug. 2002.

[46] S. Kaiser and W. A. Krzymien, "Performance effects of the uplink asynchronism in a spread spectrum multi-carrier multiple access system," *European Trans. Telecommun.*, vol. 10, no. 4, pp. 399–406, July 1999.

[47] B. Moore, *An Introduction to the Psychology of Hearing.* Brill, 2012.

[48] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[49] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*, May 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14106

[50] C. M. Zannini, A. Cirillo, R. Parisi, and A. Uncini, "Improved tdoa disambiguation techniques for sound source localization in reverberant environments," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on.* IEEE, 2010, pp. 2666–2669.

[51] D. Bechler and K. Kroschel, "Three different reliability criteria for time delay estimates," in *Proc. Eur. Signal Process. Conf. (EUSIOPCO)*, Sept. 2004, pp. 1987–1990.

[52] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, Jul. 2000.

[53] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, Sep 2003.

[54] X. Zhong, A. Mohammadi, W. Wang, A. B. Premkumar, and A. Asif, "Acoustic source tracking in a reverberant environment using a pairwise synchronous microphone network," in *Information Fusion (FUSION), 2013 16th International Conference on*. IEEE, 2013, pp. 953–960.

[55] C.-E. Chen, H. Wang, A. Ali, F. Lorenzelli, R. E. Hudson, and K. Yao, "Particle filtering approach to localization and tracking of a moving acoustic source in a reverberant room," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 4. IEEE, 2006, pp. IV–IV.

[56] E. A. Lehmann and R. C. Williamson, "Particle filter design using importance sampling for acoustic source localisation and tracking in reverberant environments," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, pp. 1–9, 2006.

[57] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[58] X. Zhong and J. R. Hopgood, "A time-frequency masking based random finite set particle filtering method for multiple acoustic source detection and track-

ing," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2356–2370, Dec 2015.

[59] J. Choi, J. Kim, and N. S. Kim, "Robust time-delay estimation for acoustic indoor localization in reverberant environments," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 226–230, Feb 2017.

요 약

최근 들어 다양한 휴대기기의 보급이 늘어남에 따라 사용자 맞춤형 서비스가 늘어나고 있다. 특히, 위치 정보에 기반한 서비스가 늘어나면서 다양한 위치 추정 기술이 주목받고 있다. 위치 기반 서비스 중에서는 GPS를 기반으로 하는 네비게이션과 같은 위치 기반 안내 서비스, 증강현실 서비스 등이 널리 쓰이면서 사용자들에게 편의를 제공하고 있다. 하지만 현재 가장 많이 쓰이는 GPS 기술은 실외에서만 사용할 수 있다는 제한이 있다. IoT, 사용자 맞춤형 로봇과 같은 실내 서비스가 많이 만들어지면서 실내에서의 위치 기반 서비스의 필요성이 대두되고 있다. WiFi, 블루투스, RFID 혹은 광센서와 같은 다양한 센서를 이용한 실내 위치 추정 기술에 대한 연구가 이루어지고 있지만 장애물, 벽으로 인하여 시야가 막혀있으면 동작하지 않거나, 센서의 특성으로 인하여 측위의 정확도가 낮다는 단점이 존재한다.

위치 추정을 위한 여러가지 센서 중 우리는 음향 센서에 주목하였다. 음향 센서, 즉 마이크를 이용하여 위치 추정을 하게 되면, 실내에 미리 설치되어 있는 스피커를 사용할 수 있다는 점과 다양한 휴대기기에 장착된 스피커 및 마이크를 이용할 수 있다는 점으로 인하여 추가적인 장비가 필요하지 않다는 장점이 있다. 또한 음향 신호는 직선거리가 장애물로 막혀있는 경우에도 수신할 수 있다는 이점이 있다.

본 논문에서는 이러한 장점을 최대한 살릴 수 있는 음향 신호를 이용한 위치 추정 시스템을 제안하였다. 음향신호를 이용하는 위치 추정 시스템은 위치를 알고 있는 스피커를 통하여 인공적으로 만든 음향 신호를 재생하고, 위치 추정의 대상인 마이크가

재생되는 신호를 녹음하여 위치를 추정하는 시스템이다. 특히, 우리는 연구를 진행하면서, 사람의 활동에 영향을 주지 않기 위하여 비가청 주파수 대역을 이용하는 신호를 설계하여 사용하였다. 위치 추정을 하기 위해서는 잔향이나 반사에 의한 신호가 아닌 직접 신호를 추정하는 것이 중요하며, 우리는 잔향이 있고 넓은 공간에서도 직접 신호를 추정하여 위치 추정의 정확도를 올리는 방법에 대하여 다음의 내용들을 제안하였다.

먼저, 넓고 잔향이 있 공간에서의 위치 추정에 적합한 음향 신호의 구조를 제안하였다. 넓은 공간에서는 원근효과(near-far effect)가 존재한다. 이는 원하는 신호가 멀리 있을 때, 가까이 존재하는 다른 신호가 원하는 신호의 수신을 방해하는 것을 의미한다. 통신에서는 송수신단의 상호작용을 통하여 이를 보완 할 수 있지만, 음향 신호를 이용하는 경우에는 송수신단이 분리 되어 있기 때문에 어려움이 있다. 우리는 통신에서 사용하는 OFDMA-CDM이라는 형태와 유사한 방식으로 신호를 만들어서 원근 효과로 인한 문제를 해결하였다. 잔향이 많은 환경에서는 직접 신호가 잔향 혹은 반사음보다 그 크기가 작아서 검출이 어려운 경우가 존재하며, 이를 위하여 직접 신호를 추정하는 방법을 제안하였다.

두번째로, 잔향이 심한 환경에서의 위치 추정 방법을 제안하였다. 잔향이 심해짐에 따라 원하지 않는 신호와 직접신호를 구분하는 것이 더 어려워지게 된다. 우리는 위치를 추정하기 위한 각 신호들의 집합을 만들고, 각각의 신뢰도를 계산하여, 가장 신뢰도가 높은 신호들을 이용하여 위치를 추정하는 방법을 제안하였다. 이 방법을 사용하면 잔향이 있거나, 직접 신호가 장애물 등으로 인해 세기가 약해진 경우에도 위치를 계산할 수 있게 된다.

마지막으로, 잔향이 심한 환경에서 비교적 적은 계산량으로 위치를 추정하는 방법을 제안하였다. 이 방법은 파티클 필터를 기반으로 동작하며, 이는 가중치를 가진 입자들을 이용하여 위치를 추정하는 방법이다. 입자들의 가중치는 관측값으로부터 우도를 계산하여 얻어지며, 우리는 관측값으로 부터 위치계산에 적합한 우도를 계산하는 방법을 제안하였다. 각 소스의 관측값에서 신뢰도를 우도에 반영하는 방법, 효율적이

지만 정확도를 올리는 계산 방법을 제안하였고, 또한 관측값에서 직접 신호가 포함된 부분을 추정하여 위치 계산에 반영하는 방법을 제안하였다. 이러한 방법을 통하여, 잔향이 심한 경우에도 파티클 필터를 이용하여 반복적으로 계산하여 정확한 위치 추정이 가능하다는 것을 확인하였다.

이와 같은 실내 환경에서의 위치 추정을 위하여 제안한 다양한 방법들은 잔향 시간을 바꿔가며 구성한 시뮬레이션 환경에서 성능을 확인하였다. 여러 가지 파라미터를 바꿔가면서 제안한 방법의 성능을 확인하고, 기존 논문들에서 사용한 데이터 구조 및 측위 방법과 성능을 비교하였고 제안한 방법의 우수성을 확인하였다. 또한 넓고 잔향이 심한 실제 환경에서도 실험을 진행하였다. 일련의 실험을 통하여 제안한 위치 추정 시스템이 실제 환경에서 마이크의 위치를 추정하는 방법으로 적합하다는 것을 확인할 수 있었다.

주요어: 위치 기반 서비스, 실내 위치 추정, 마이크의 위치 추청, 시간 지연 추정, 잔향 환경, 파티클 필터

학 번: 2010-20905