Acoustic Positioning Using Multiple Microphone Arrays

Hui Liu
Evangelos Milios

Technical Report CS-2004-01

January 23, 2004

Faculty of Computer Science
6050 University Ave., Halifax, Nova Scotia, B3H 1W5, Canada

Acoustic Positioning Using Multiple Microphone Arrays

Hui Liu, Evangelos Milios
Faculty of Computer Science
Dalhousie University
Halifax, Canada B3H 1W5
{hliu, eem}@cs.dal.ca

January 23, 2004

Abstract

An important problem associated with the navigation of an autonomous vehicle is its accurate localization. We are investigating acoustic techniques to solve the localization problem. In the case of an autonomous underwater robot, we use an array of hydrophones floating at the surface, whose absolute position can be measured via GPS. We estimate the direction of arrival of sound at the hydrophone array from a source emitting sound in the audio range that is mounted on the robot. Estimation of the direction of arrival at multiple hydrophone arrays allows the positioning of the underwater robot in 3D space. Novel array designs are considered, and an optimal position estimation approach in the presence of measurement errors is provided. Experimental results with real data are provided both in the air (using arrays of condenser microphones) and in the water.

Contents

List of Tables

List of Figures

Acknowledgements

The research was supported by a grant from the The Institute for Robotics and Intelligent Systems (IRIS), a federally funded Network of Centres of Excellence managed by PRECARN Inc., and by the Natural Sciences and Engineering Research Council of Canada.

I would like to thank Dr. Milios, not only for supervising this thesis, but also for giving me advice, encouragement, and support over the last two years.

I also would like to thank my thesis defense committee members, Dr. Gao and Dr. Gu, whose comments substantially improved the thesis.

Special thanks to Yael Kollet, Kori MacCara, Lingyan Zhang, Liwen Zhou, and Ye Liang, for their support and help with the logistics of the experiments. Bill Kapralos of York University provided me his Master's thesis code, and helped me in adapting it to my project.

Shawn Fraser, Pat McGrath and their staff gave us generous access to the Dalhousie University pool facility.

Chapter 1

Introduction

Modern oceanographic studies have created a demand for precise underwater navigation and positioning systems. The most practical means of achieving precise localization under the ocean surface is through the use of acoustic techniques. Acoustic transponders for underwater localization usually work at ultrasonic frequency range and require expensive special purpose signal processing hardware.

Acoustic positioning is also useful in the context of teleconferencing [5]. Existing appliances use multiple microphones to localize a speaker in the room and steer a video camera to that speaker.

Rabinkin [13] proposes an algorithm using arrays of omnidirectional microphones to localize a sound source with beamforming and generalized time-delay techniques. However, it requires simultaneous and synchronized recording and digitization from all microphones. Therefore, in the underwater case, Rabinkin's method in its original form is not very practical because it requires the synchronous A/D conversion from multiple arrays mounted on buoys that are tens of meters apart, with the only practical communication between them being by wireless ethernet. Moreover, Rabinkin's algorithm is based on iterative non-linear optimization, therefore it requires an initial source position to start the iteration.

In this report, we intend to precisely estimate the direction of arrival from a sound-emitting object to a passive local microphone array. We use multiple microphone arrays in the audio frequency range at known positions to measure the directions of arrival of sound based on the far-field assumption. The sound source position is estimated at the intersection of the multiple direction lines in 3D space.

The motivation for our research is underwater vehicle localization. In this context, a sound source is mounted on the robot, and omnidirectional hydrophone arrays are floating near the surface, whose locations can be determined by Global Positioning System (GPS). Differential GPS can achieve a position accuracy of $30cm$ [4], which is adequate for the

absolute localization of hydrophone arrays in an open water experiment. We then locate the robot by estimating the position of its sound source with respect to the hydrophone arrays, as shown in Figure 1.1.

The contributions of our method are the following: (a) it is a closed-form solution for the localization of source position. (b) it is decentralized, in that each array performs its own A/D conversion and direction of arrival estimation in an asynchronous manner with respect to the other arrays. Once direction vectors have been computed, they can easily be communicated to the other arrays for localization, via wireless ethernet. Another advantage of decentralized processing is that we can add more arrays to improve the area coverage or accuracy of the localization without any problem. In Rabinkin's method, we would need ever more sophisticated centralized digitization hardware to accommodate additional microphones, and ever longer cables to transmit the analog sound to the digitization hardware.

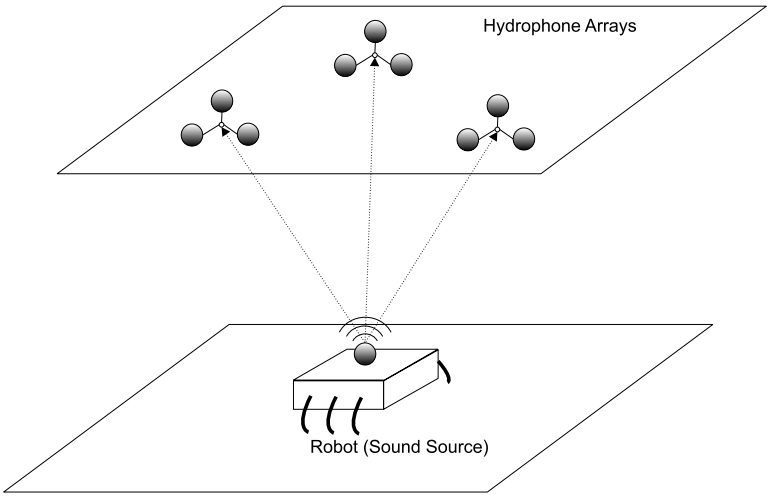

Figure 1.1: Multiple hydrophone arrays floating on the surface with each estimating a direction of arrival of sound from a sound source mounted on the robot. The position of robot is at the intersection of direction lines.

The report is structured as follows. A review of existing acoustic systems is presented in Chapter 2. The methodology and localization principle are discussed in Chapter 3. Chapter 4 describes the signal processing mechanism for time delay estimation. The results of the simulation, and of the air and the pool experiments are presented in Chapter 5. Finally, conclusions and directions for future research are presented in Chapter 6.

Chapter 2

Background

Various acoustic methods have been described in the literature for the localization of air and ground objects, and for underwater applications. The military has been using towed arrays for a long time for detection and localization of submarines [1].

One typical acoustic positioning application, which finds wide use in teleconferencing, is to localize a speaker based on her voice [13] [5] [14]. Reid [14] estimates the direction of arrival of sound by an omnidirectional microphone pair. The microphone pair is mounted on a pan-tilt unit with a fixed baseline and two rotational degrees of freedom. The sound direction is estimated by time delays between the microphones in different microphone pair positions. Therefore, a persistent signal is required for the direction estimation. Kapralos [5] reports a position estimation method with an array of four omnidirectional microphones mounted in a static pyramidal shape. This method uses delay and sum beamforming to determine the direction of arrival of the sound. In the absence of source direction information, it searches by focusing the beamformer to every possible source direction. Audio is combined with an panoramic camera.

Another acoustic method proposed by Rabinkin [13] uses an array of a large number of omnidirectional microphones to localize the speaker in a teleconferencing room. The algorithm estimates the time delay of arrival of sound between all microphone pairs in the array, and then constructs an least-mean-squares error function in terms of the unknown source position (x, y, z), and the measured time delays. Source position is estimated by minimizing the error function, which involves non-linear optimization.

The formulation of Rabinkin's approach is the following. Consider a sound source $\mathbf{s}$ with unknown position $[x, y, z]$ and a microphone pair $\mathbf{M_i}$, $\mathbf{M_j}$, with coordinates $\mathbf{M_i} = [M_{ix}, M_{iy}, M_{iz}]$ and $\mathbf{M_j} = [M_{jx}, M_{jy}, M_{jz}]$, as shown in Figure 2.1. It takes time τ_i for the sound signal to propagate from $\mathbf{s}$ to $\mathbf{M_i}$, and τ_j from $\mathbf{s}$ to $\mathbf{M_j}$. The propagation time can be obtained by the following expression:

$$\tau_i = \frac{d_i}{c} = \frac{\sqrt{(M_{ix}-x)^2+(M_{iy}-y)^2+(M_{iz}-z)^2}}{c} \tag{2.1}$$

where d_i is the distance between $\mathbf{s}$ and $\mathbf{M_i}$, and c is the speed of sound.

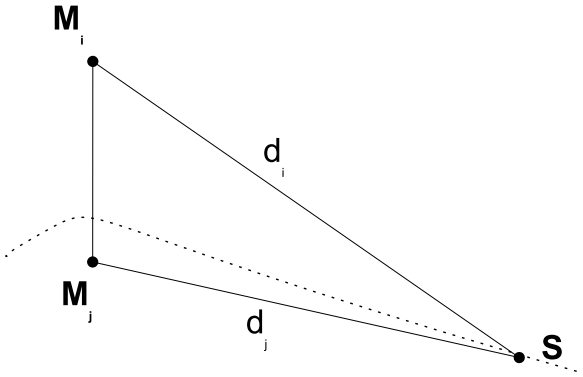

Figure 2.1: The distances from the sound source $\mathbf{s}$ to microphone pair $\mathbf{M_i}$ and $\mathbf{M_j}$ are d_i and d_j respectively.

The propagation time difference D_{ij} corresponds to the distance difference between d_i and d_j:

$$D_{ij} = \tau_i - \tau_j = \frac{d_i - d_j}{c} \tag{2.2}$$

Therefore, a given set of microphone pairs leads to an associated set of equations constraining the position of the source:

$$\begin{cases} \frac{d_1-d_2}{c} = D_{12} \\ \cdots \\ \frac{d_i-d_j}{c} = D_{ij} \\ \cdots \end{cases} \tag{2.3}$$

Thus, a set of three equations uniquely specifies the source position. For a set of four or more microphones, the following minimization function is used to solve for the source position:

$$\min_{x,y,z} \Big(\sum_{i,j} (\frac{d_i - d_j}{c} - D_{ij})^2 \Big) \tag{2.4}$$

Rabinkin uses a numerical method to solve this minimization problem. An initial estimate of the source position is required to start the search. Experimental results reported in [13] involve 20 different source locations regularly distributed in a 4×5 matrix towards an array with eight microphones. In one room setting the accuracy of the estimates reported was within $30cm$. However, the source location performance with the other room setting was

poor, and it often failed to find a correct estimate for the source. The probable cause was the considerable amount of reverberation in the room, which made the time delay estimation algorithm less reliable. We also tested this algorithm with the same settings of the air experiments. The quality of the results of our method are comparable to Rabinkin's. However, Rabinkin's method requires an initial estimate of the source position for the numerical minimization, since a closed-form solution of Eq. (2.3) is not available. Another problem of this approach is that it requires simultaneous recording and digitization of the signals from all microphones. Therefore, special purpose multichannel A/D hardware and long cables are required to pull the analog signals together for digitization. If the microphone arrays are spatially separated, long cables are problematic.

Our approach is to group the microphones into different arrays, and decentralize the digitization and processing, so that each array independently computes direction of arrival, and different arrays exchange direction of arrival estimates for source localization. We measure the time delay between each microphone pair and apply Eq. (2.3) to the microphones within the same array. Then the minimization function of Eq. (2.4) is "decentralized" to the following form:

$$\min_{x,y,z} \Big(\sum_{k} \sum_{i,j} (\frac{d_i - d_j}{c} - D_{ij})^2 \Big) \tag{2.5}$$

where k is the number of microphone array.

Autonomous underwater vehicles (AUV) are receiving increasing interest from marine scientists who use them to perform tasks of underwater object inspection, geophysical field measurements and environmental assessment and monitoring. A challenging issue is AUV localization and navigation. Traditional vision techniques for mobile robot positioning are difficult to use underwater. GPS (Global Positioning System) signals are not available underwater. Therefore, acoustic positioning is the method of choice, especially in the applications involving long range measurement.

Currently, commercial acoustic positioning systems are built upon the concepts of long baseline system (LBL) [18][6][9], short baseline system (SBL) [19][7] and ultra short baseline system (USBL) [11]. These systems have different methods for localization and most of them require special purpose signal transponder and signal processing apparatus.

A conventional mode of SBL system uses three hydrophones that are mounted on a vessel bottom and a beacon is placed on the sub-sea vehicle or equipment [7]. The travel times of the acoustic signal from the beacon to the hydrophones are measured by interrogating the beacon, which responds by transmitting a signal, and the ranges from the beacon to the hydrophones are the signal propagation times multiplied by the underwater sound of speed. The position of the beacon is the intersection of three spheres with radii being the ranges

measured. A simulation program is described in [7] for testing this approach.

A USBL positioning method is addressed by Opderbecke [11]. The method uses a surface-ship-mounted acoustic array consisting of four hydrophones to measure range from time-of-flight and direction from phase difference of an acoustic signal emitted by a transponder beacon on the underwater vehicle. The central problem of all methods that use time-of-flight, which is solved by interrogating the beacon who must be able to both receive and transmit signals, is how to synchronize the signals during the measurement.

Klimley [6] describes an LBL method, where three hydrophones on sonobuoys are aligned as a triangular array in an area of $1km^2$. The delays between the time of arrival of the same signal pulse at the nearest buoy and the two more distant buoys are calculated to construct two hyperbolas together with the distances between the sonobuoys and speed of sound. The position of the transmitter is estimated as the intersection of hyperbolas. This system has been tested in a simulation program, demonstrating a standard deviation in the position estimate of less than $10m$. However, the positions of the transmitters are limited to the area within the triangular sonobuoy configuration.

Towed arrays and sonobuoy arrays are used in underwater surveillance systems for submarine detection and localization. Typical towed array processing techniques include adaptive beamforming, adaptive cancellation of tow vessel noise, bearing estimation using eigenvectors, and ranging techniques such as triangulation or wave-front curvature [2]. And usually the implementation of these techniques require sophisticated and expensive signal processing hardware.

Finally, passive acoustic techniques in the underwater domain are important in the study of marine life [17]. Detection and localization of large marine mammals using passive acoustics is an essential component of their protection action as endangered species [10, 16].

Chapter 3

Sound Source Localization

We now present the geometric model for the sound source localization in this chapter. The essence of the model is to first estimate the direction of the arrival of sound at each microphone array based on the far-field assumption, and then to determine the sound source position as the intersection of direction lines towards the source out of the multiple microphone arrays.

We have also considered other geometric models in our work. Further details about the methodology and the simulation results of those models are presented in Appendix A.

3.1 Measurement of Direction of Arrival of Sound

The measurement of the direction of arrival of sound is based on the far-field assumption, that requires the sound source to be sufficiently far from the microphones for the acoustic wave reaching the microphones to be approximately planar.

Consider two omnidirectional microphones $\mathbf{M_1}$, $\mathbf{M_2}$ with coordinates $\mathbf{M_1} = [M_{1x}, M_{1y}, M_{1z}]$ and $\mathbf{M_2} = [M_{2x}, M_{2y}, M_{2z}]$ as shown in Figure 3.1. The time difference for the sound to propagate from $\mathbf{M_1}$ to $\mathbf{M_2}$ is t_{12}, which is related to the angle of arrival γ and can be calculated by the constraint:

$$\cos\gamma = \frac{ct_{12}}{|\mathbf{b_{12}}|} = \frac{cn_{12}}{f|\mathbf{b_{12}}|} \tag{3.1}$$

where c is the speed of sound, $|\mathbf{b_{12}}|$ is the length of the baseline between $\mathbf{M_1}$ and $\mathbf{M_2}$, n_{12} is time delay in samples, f is the sampling frequency.

Consider the unknown direction from the reference point of the microphone array to the source as a unit vector $\mathbf{s}[s_x, s_y, s_z]$. The relation between $\mathbf{s}$ and the baseline vector $\mathbf{b_{12}}[b_{12x}, b_{12y}, b_{12z}] = \mathbf{M_1} - \mathbf{M_2} = [M_{1x} - M_{2x}, M_{1y} - M_{2y}, M_{1z} - M_{2z}]$, where γ is the

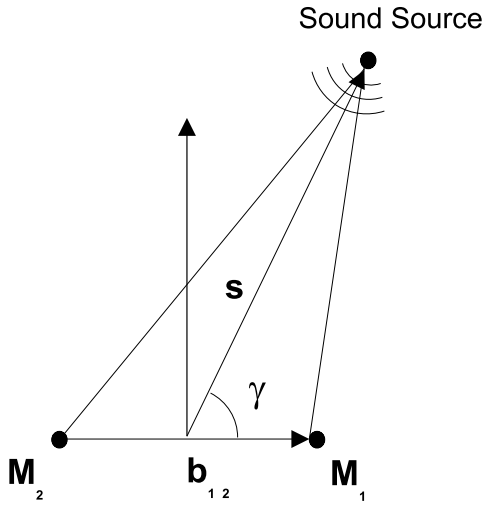

Figure 3.1: Two microphones $\mathbf{M_1}$ and $\mathbf{M_2}$ form a baseline $\mathbf{b_{12}}$, listening to the same source. The direction of sound event $\mathbf{s}$ arrives at the microphones with angle γ.

incidence angle of the sound wave onto the microphone array, is:

$$\mathbf{s} \cdot \mathbf{b_{12}} = |\mathbf{b_{12}}| \cos\gamma \tag{3.2}$$

or equivalently,

$$s_x b_{12x} + s_y b_{12y} + s_z b_{12z} = |\mathbf{b_{12}}| \cos\gamma \tag{3.3}$$

Combining Eq. (3.1) and Eq. (3.3) we obtain:

$$s_x b_{12x} + s_y b_{12y} + s_z b_{12z} = c t_{12} \tag{3.4}$$

Thus a set of three linear equations can be solved to uniquely determine the three unknown variables s_x, s_y and s_z. This implies that three microphones are required in one microphone array for finding the sound direction. For arrays with more than three microphones, we have an over-determined problem, for which an optimal solution for $\mathbf{s}$ can be computed via a least mean squares approach. The error $f_1(s_x, s_y, s_z)$ in satisfying Eq. (3.4) is:

$$f_1(s_x, s_y, s_z) = s_x b_{12x} + s_y b_{12y} + s_z b_{12z} - c t_{12} \tag{3.5}$$

There is also an implicit nonlinear constraint that $s_x^2 + s_y^2 + s_z^2 = 1$ since $\mathbf{s}$ is a unit vector:

$$f_{nonlinear}(s_x, s_y, s_z) = s_x^2 + s_y^2 + s_z^2 - 1 \tag{3.6}$$

Therefore, the solution for the sound direction can be obtained by solving the nonlinear

minimization problem shown in Eq. (3.7). An iterative non-linear optimization algorithm is used to solve this problem [12], where the initial solution for starting the search is obtained by solving the linear least squares problem obtained by excluding the nonlinear constraint.

$$\min_{s_x,s_y,s_z}(f_{nonlinear}^2 + \sum_i f_i^2) \tag{3.7}$$

3.2 Far Field Assumption Error

As we described in the previous section, the estimation of the direction of the arrival of the sound is based on the far-field assumption. We now attempt to quantify the error this assumption introduces. The incidence angle $\alpha = \frac{\pi}{2} - \gamma$ shown in Figure 3.2 can be calculated through the constraint in Eq. (3.1).

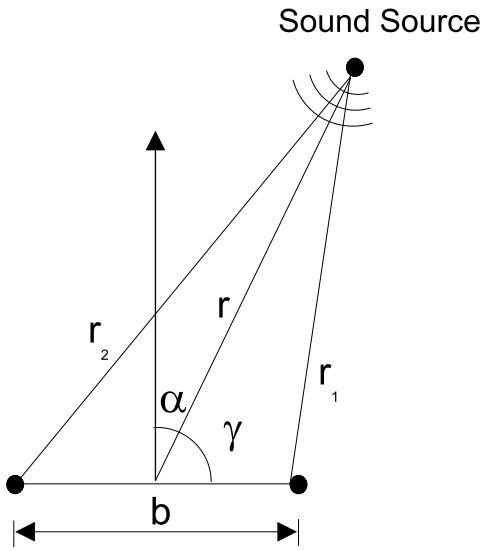

Figure 3.2: Geometry of incidence of sound on two microphones. Angle of incidence α. Length of microphone baseline is b. The distances from the sound source to the two microphones, and to the center of the baseline are r_1, r_2, and r respectively.

The error between the actual angle α and its far-field approximation (Fig. 3.2) is the following:

$$Error = |\alpha - \alpha_{approx}| = |\alpha - \arcsin(\frac{r_2 - r_1}{b})| \tag{3.8}$$

where b is the length of the baseline between the two microphones and r is the distance from the sound source to the center of the baseline. r_2 and r_1 are distances to the two

microphones. They are dependent on b, r, γ through the cosine law:

$$\begin{aligned} r_1 &= \sqrt{r^2 + \frac{b^2}{4} - rb\cos(\gamma)} \\ r_2 &= \sqrt{r^2 + \frac{b^2}{4} + rb\cos(\gamma)} \end{aligned} \tag{3.9}$$

Therefore, the approximation error in Eq. (3.8) can be transformed into a function of angle α and the ratio r/b:

$$Error(\alpha, \frac{r}{b}) = |\alpha - \arcsin(\sqrt{(\frac{r}{b})^2 + \frac{r}{b}\sin\alpha + \frac{1}{4}} - \sqrt{(\frac{r}{b})^2 - \frac{r}{b}\sin\alpha + \frac{1}{4}})| \tag{3.10}$$

Figure 3.3 shows the far-field assumption error over the full range of α. It is noted that when the ratio of r/b is greater than 3, the error is less than 0.4^o.

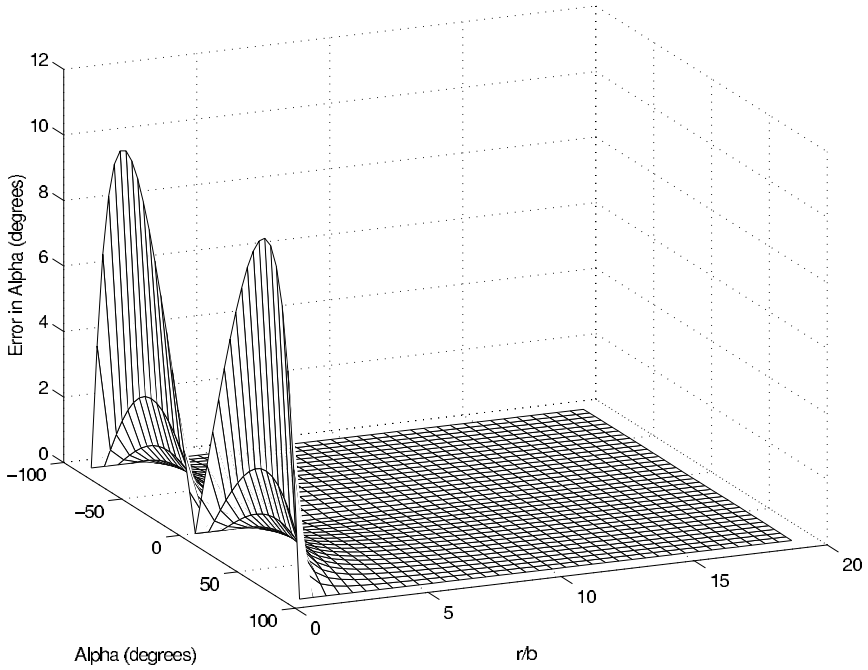

Figure 3.3: The plot of the error due to the far-field assumption, as a function of the incidence angle α and the ratio of the distance to the source and the distance between two microphones $\frac{r}{b}$.

3.3 Estimation of Sound Source Position

The position of the sound source is estimated as the intersection of multiple lines in three-dimensional space. Each line is defined by the reference point of a microphone array and

the direction vector from the microphone array to the source. Due to measurement noise in the direction vectors and the microphone positions, these lines will not, in general, intersect exactly at a single point. Therefore, the optimal estimate of the source position is the point that has minimal overall distance from these lines.

Consider the center (reference point) $\mathbf{P_i}$ of a microphone array and the associated direction of arrival of sound $\mathbf{s_i}$. A line l_i in three-dimensional space can be defined by its start at $\mathbf{P_i}$ and its direction $\mathbf{s_i}$, as shown in Fig. 3.4. Every point $\mathbf{P}$ lying on l_i satisfies the following constraint for some real number λ:

$$\mathbf{P} = \mathbf{P_i} + \lambda \mathbf{s_i} \tag{3.11}$$

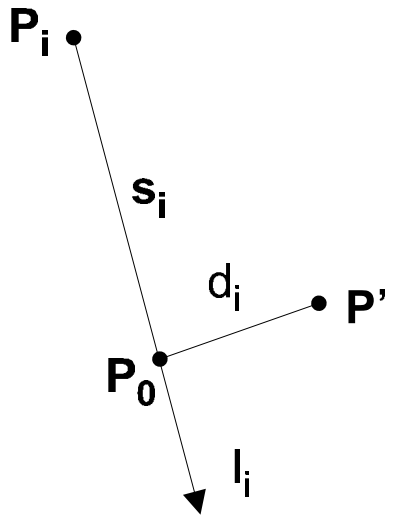

Figure 3.4: The center of a microphone array $\mathbf{P_i}$ and estimated direction from the microphone array to the source $\mathbf{s_i}$ form a line l_i. d_i is the distance from the sound source position $\mathbf{P'}$ to l_i, representing the error in the estimate of $\mathbf{s_i}$, as the line should ideally contain $\mathbf{P'}$.

Consider the unknown sound source position $\mathbf{P'}[x, y, z]$. It has the following relation to the direction vector $\mathbf{s_i}$:

$$(\mathbf{P'} - \mathbf{P_0}) \cdot \mathbf{s_i} = 0 \tag{3.12}$$

where $\mathbf{P_0}$ is the projection of $\mathbf{P'}$ onto l_i. As a point of l_i, $\mathbf{P_0}$ also satisfies relation Eq. (3.11), where λ is a parameter to be determined:

$$\mathbf{P_0} = \mathbf{P_i} + \lambda \mathbf{s_i} \tag{3.13}$$

If we combine Eq. (3.12) and Eq. (3.13), and eliminate $\mathbf{P_0}$, then we can solve for λ in terms

of $\mathbf{P'}$, $\mathbf{P_i}$ and $\mathbf{s_i}$:

$$\lambda = (\mathbf{P'} - \mathbf{P_i}) \cdot \mathbf{s_i} \tag{3.14}$$

By applying Eq. (3.14) to Eq. (3.13) we obtain:

$$\mathbf{P_0} = \mathbf{P_i} + \Big((\mathbf{P'} - \mathbf{P_i}) \cdot \mathbf{s_i}\Big)\mathbf{s_i} \tag{3.15}$$

The distance d_i between $\mathbf{P'}$ and $\mathbf{P_0}$, which is the distance from $\mathbf{P'}$ to l_i, can be expressed as:

$$d_i = |\mathbf{P'} - \mathbf{P_0}| \tag{3.16}$$

or equivalently

$$d_i^2 = (\mathbf{P'} - \mathbf{P_0}) \cdot (\mathbf{P'} - \mathbf{P_0}) \tag{3.17}$$

By applying Eq. (3.15) to Eq. (3.17), d_i^2 can be obtained by the following constraint with respect to $\mathbf{P'}$, $\mathbf{P_i}$ and $\mathbf{s_i}$:

$$d_i^2 = \Big\|\Big(\mathbf{P'} - \mathbf{P_i}\Big) - \Big((\mathbf{P'} - \mathbf{P_i}) \cdot \mathbf{s_i}\Big)\mathbf{s_i}\Big\|^2 \tag{3.18}$$

In terms of the (unknown) coordinates x, y, z of the sound source position $\mathbf{P'}$ Eq. (3.18) has the form:

$$\begin{aligned} d_i^2 \;=\; & c_{i1}x^2 + c_{i2}y^2 + c_{i3}z^2 + \\ & c_{i4}xy + c_{i5}xz + c_{i6}yz + \\ & c_{i7}x + c_{i8}y + c_{i9}z + c_{i10} \end{aligned} \tag{3.19}$$

where $c_{i1} \ldots c_{i10}$ are constants determined by $\mathbf{P_i}$ and $\mathbf{s_i}$. Therefore, given a set of microphone arrays, the solution for the sound source position coordinates x, y, z can be computed through the minimization equation:

$$\min_{x,y,z} \Big(\sum_i d_i^2 \Big) \tag{3.20}$$

where the summation extends over all microphone arrays i:

$$\begin{aligned} \min_{x,y,z}(\quad & \sum_i c_{i1}x^2 + \sum_i c_{i2}y^2 + \sum_i c_{i3}z^2 + \\ & + \sum_i c_{i4}xy + \sum_i c_{i5}xz + \sum_i c_{i6}yz + \\ & + \sum_i c_{i7}x + \sum_i c_{i8}y + \sum_i c_{i9}z + \sum_i c_{i10}) \end{aligned} \tag{3.21}$$

To solve the minimization problem, we set the first partial derivatives of Eq. (3.21) over x,

y and z to 0:

$$\begin{cases} 2\sum_i c_{i1}x + \sum_i c_{i4}y + \sum_i c_{i5}z + \sum_i c_{i7} = 0 \\ \sum_i c_{i4}x + 2\sum_i c_{i2}y + \sum_i c_{i6}z + \sum_i c_{i8} = 0 \\ \sum_i c_{i5}x + \sum_i c_{i6}y + 2\sum_i c_{i3}z + \sum_i c_{i9} = 0 \end{cases} \tag{3.22}$$

To determine whether the solution to the above system of linear equations is a minimum, we inspect the matrix of second partial derivatives of Eq. (3.21) over x, y and z:

$$\begin{bmatrix} f''_{xx} & f''_{xy} & f''_{xz} \\ f''_{yx} & f''_{yy} & f''_{yz} \\ f''_{zx} & f''_{yz} & f''_{zz} \end{bmatrix} \tag{3.23}$$

or, equivalently:

$$\begin{bmatrix} 2\sum_i c_{i1} & \sum_i c_{i4} & \sum_i c_{i5} \\ \sum_i c_{i4} & 2\sum_i c_{i2} & \sum_i c_{i6} \\ \sum_i c_{i5} & \sum_i c_{i6} & 2\sum_i c_{i3} \end{bmatrix} \tag{3.24}$$

If the determinant of Matrix (3.24) is greater than 0, then the solution of the set of linear equations (3.22) is a minimum of Eq. (3.21).

Chapter 4

Signal Processing for Time Delay Estimation

In order to calculate reliable time delay between the arrivals of the sound signals, we correlate two channels of a window of audio data from two microphones, and look for the peak of the correlation function. The location of the peak corresponds to the time-delay estimate. Before correlation, filtering is carried out to reduce noise, and a signal level test is performed to check for the presence of a sound source. In this work, we follow the same signal processing steps that are described in [14]. The flowchart in Figure 4.1 is a summary of the processing on the sampled audio signals for time-delay estimation.

4.1 Filtering

Most human-audible sounds lie in the frequency range from a few Hz to about $20KHz$. A $60Hz$ electric interference is common in audio signals. We also want to cut off high frequency noise that is above $4000Hz$. Therefore, frequencies that are not in the region of interest (200 - $4000Hz$) will be filtered out. This is accomplished through the use of a bandpass digital FIR filter, which is described in [5].

4.2 Signal Level Test

A test is performed to determine the presence of signal. The simplest approach is based on a comparison of the signal magnitude E_{signal} to the magnitude of the background noise. If the signal is weak relative to background, the above magnitude level test may not be sufficient. A test based on *signal variance* was introduced in [5]. Signal variance captures the overall

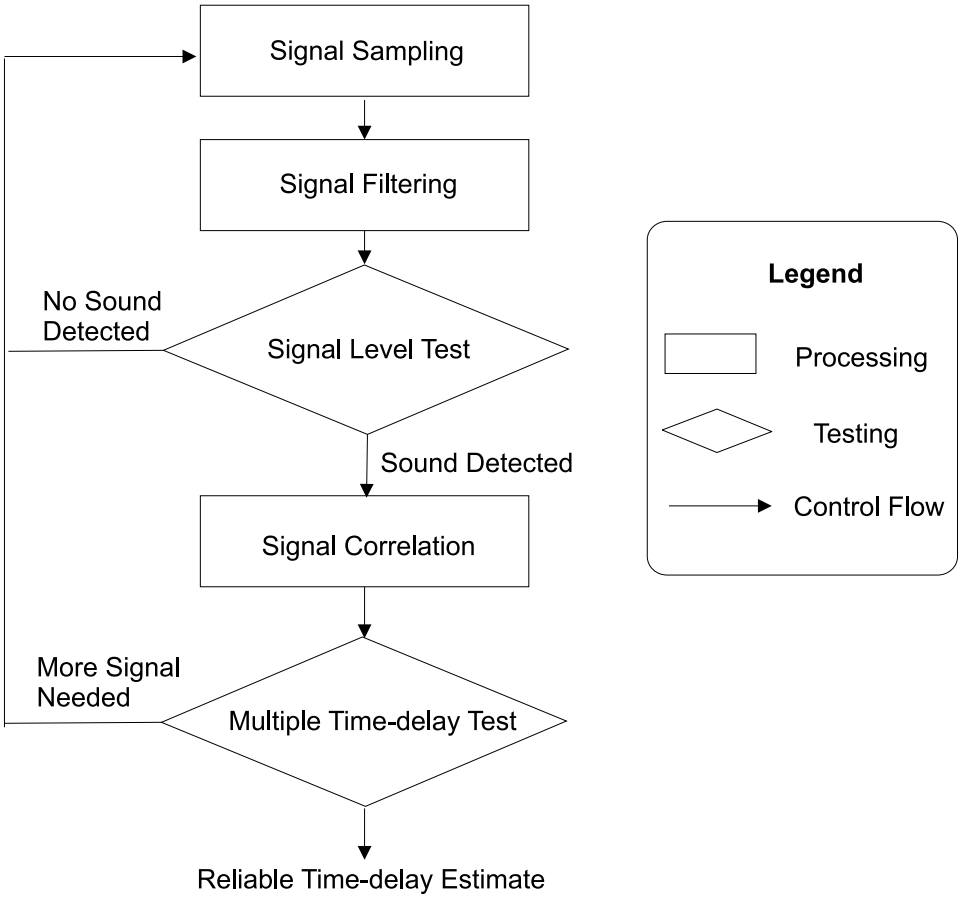

Figure 4.1: Signal processing flowchart for time delay estimation

deviation of signal samples from the mean. Signal variance may be computed with using either Eq. (4.1) or Eq. (4.2).

$$\overline{s} = \frac{1}{n}\sum_{j=0}^{n-1} |s[j]| \tag{4.1}$$

$$V_{signal} = \frac{1}{n}\sum_{j=0}^{n-1} (s[j] - \overline{s})^2 \tag{4.2}$$

Generally, the variance of a signal is greater than the variance of the background noise, as shown in Figure 4.2.

In our work, we use a threshold value V_{thresh} to determine the presence of sound. The maximum variance of the background noise $\max V_{bckgrnd}$ is dynamically measured, and V_{thresh} is computed by Eq. (4.3), where λ is a constant real number. The value of λ is determined experimentally.

$$V_{thresh} = (\max V_{bckgrnd}) + \lambda \tag{4.3}$$

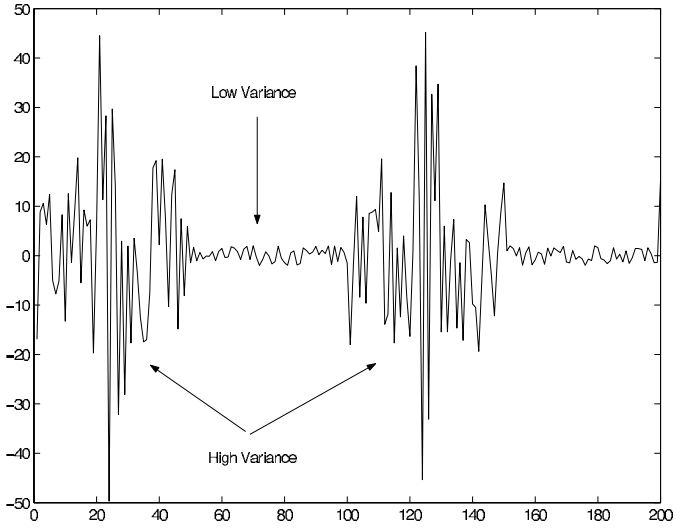

Figure 4.2: Graphical illustration of low and high variance. Comparison of variance is used to detect the presence of sound.

We should note that if we have additional knowledge of the signal source, for example, when we attach a known source onto an underwater vehicle for the purpose of localizing it, then a more powerful matched filtering approach can be applied for signal detection.

4.3 Signal Correlation

The correlation function, which we use in this work, is given by:

$$f(i) = \sum_{t} s_1(t+i)s_2(t) \tag{4.4}$$

where s_1 and s_2 are two sound channels over the time window t and i is the amount of time shift, which is the independent variable of the correlation function. The value of i that yields the maximum output of function $f(i)$ corresponds to the time delay estimate.

The maximum possible time delay between two microphones equals the length of the baseline divided by the speed of sound. We therefore require that the time delay of the maximum peak of the correlation function be under the maximum delay.

4.4 Multiple Time Delay Test

The final step for the time delay estimation is to reject outliers by clustering the time delays estimated from a number of consecutive and non overlapping signal time windows. We

discard outliers and compute the mean value over the remaining as the final time delay estimate.

Agglomerative clustering [3, 8] is used. Consider we have five time delay estimates $\{t_1, t_2, t_3, t_4, t_5\}$. We initially place each number into its own group. Therefore, we start with five groups. Each of these groups contains only a single number and the initial centroid of each group is the number itself. Clustering proceeds as follows, based on a carefully selected threshold distance.

1. Compare all centroid pairs.
2. Compare the distance between the closest centroid pair to the threshold value.
   (a) If the distance between this closest pair is less than the threshold distance, then these two groups are merged into a single group and a new centroid for the new group is generated. Return to Step 1 to continue the clustering.
   (b) If the distance between the closest pair is greater than the threshold, the clustering stops.

Chapter 5

Experimental Results

In this chapter we describe our experiments. We performed three kinds of experiments:

1. Simulations for the underwater case.
2. Air experiments using microphones in an arrangement similar to that in the underwater case.
3. Water experiments in a pool using hydrophones.

In the experiments, we generated impulsive sound by hitting a metal appliance approximately once per second, and used one or more microphone arrays to estimate the location of the sound source in 3D space. The algorithm for time delay estimation is given in Figure 5.1. It uses the method described in Chapter 4 to correlate the signals received at the listening apparatus.

Issues to consider in an open water environment include the variation of the speed of sound with water temperature, the accuracy of the Global Positioning System GPS for the absolute localization of the hydrophone arrays and the movement of the arrays due to currents and waves. For this report, we carried water experiments in a pool, where these issues could be disregarded.

5.1 Simulation

A simulation was implemented on Matlab6.0 to test the positioning method described in Chapter 3, and compare it with Rabinkin's method [13]. The coordinates of the microphones are shown in Table 5.1, and the coordinates of the sound source are $[0, 0, 0]$. Figure 5.2 shows the arrangement of the hydrophones and the sound source. The speed of sound in the simulation was set to be $c = 1,482.0 m/s$.

```
Algorithm: Time delay estimation
Input: Two channels of audio signals,
       start and end index of correlation window.
Output: Time delay between two channel of signals.

Method:
(1):  repeat:
(2):      read audio data from two channels;
(3):      compute signalVariance of audio data;
(4):      if signalVariance > dynamicVarianceThreshold, then
(5):          for every timeShift in correlationWindow
(6):              compute correlationFunction;
(7):              if output > maxCorrelationOutput, then
(8):                  let output be maxOorrelationOutput;
(9):                  save this timeShift as indexOfPeak;
(10):         if indexOfPeak < precalculatedMaxDelay, then
(11):             timeDelay = indexOfPeak;
(12): until timeDelay is found;
```

Figure 5.1: Time delay estimation algorithm

In order to simulate errors that exist in the measurement of the time delay in samples between the arrival of signals, we add a random number in a specific range to the time delay estimate at the each measurement. We estimate the sound source position 25 times, and calculate the average and the standard deviation of the estimates. Then the range of the error is increased and the same process is repeated, as shown in Figure 5.3. Figure 5.4 shows the results of both our method described in Chapter 3 and Rabinkin's method [13].

5.2 Air Experiments

5.2.1 Experimental Setting

This experiment took place in a rectangular room $16m$ long, $9m$ wide, and $3m$ high. The room is carpeted and has sound-absorbent ceiling tiles. The arrangement of the room and the placement of the apparatus are shown in Figure 5.5 and in Figure 5.6.

The listening apparatus consists of two sets of microphones. Each set has four omnidirectional Genexxa 3303003 electret condenser microphones (with diameter of approximately $8mm$) mounted at the corners of a square wooden frame with side length of $0.3m$.

As described in Sec. 3.2, in order for the far-field assumption to be applicable, the ratio of distance r to length of the baseline b must be sufficiently large. In this experiment, the

(unit: m)

	x	y	z		x	y	z
M_1	-1.0	0.15	2.0	M_5	1.3	0.15	2.0
M_2	-1.3	0.15	2.0	M_6	1.0	0.15	2.0
M_3	-1.3	-0.15	2.0	M_7	1.0	-0.15	2.0
M_4	-1.0	-0.15	2.0	M_8	1.3	-0.15	2.0

Table 5.1: This table shows the microphone position arrangement used in the simulation program, where $M_1 - M_4$ form one microphone array and $M_5 - M_8$ form another array for the estimation of sound source.

sound source is placed at a minimum distance of $1.2m$ which corresponds to $r/b = 4$. The resulting far-field error is less than 0.23^o, as shown in Figure 5.7.

The main tool that we use in the measurement of ground truth is a measuring tape. To minimize measurement error, an optimal measurement method is used. Further details regarding the distance measurement and the coordinate system construction can be found in Appendix B.

We use an Echo Layla 24 Digital Audio Recording unit as the analog-to-digital (A/D) converter. Table 5.2 shows the parameter specifications for the experiments.

Parameter	**Value**
Sampling Frequency	44100Hz
Sample Resolution	16 bits
Sample Size	2048
Low Pass Filter Cutoff Frequency	4000Hz
High Pass Filter Cutoff Frequency	200Hz
Filter Coefficients	128 (for low and high pass)
$\mathbf{V_{thresh}}$	Dynamically Determined

Table 5.2: Parameter Specifications for The Air Experiments.

In our air experiments, the sound source is at a place slightly above the ground and generates an impulsive sound (hand clapping or hitting a metal appliance) approximately every $1.5s$ interval. This time interval needs to be maintained in the experiments so as to keep successive impulses and their reflections apart. The spectrogram of the sound is shown in Figure 5.8. The frequency of sound lies in the range of 200 - $4000Hz$. In addition to the impulsive sound source, there are other sound sources. For example, there are air conditioning and ventilation fans running. This background noise is not very loud and is usually below $200Hz$. Such noise is filtered out by the bandpass filtering.

Based on our setup, the range of the time delay values in terms of samples can be determined through Eq. (3.1). Considering a sound source at $\gamma = 0^o$, and with sampling frequency $f = 44100$, baseline length $= 0.3m$ and speed of sound $c = 345m/s$ Eq. (3.1) yields:

$$\pm n = \pm \frac{f|\mathbf{b}|}{c} \approx \pm 38 \tag{5.1}$$

Therefore, the number of possible time delays is in the range of $[-38, 38]$ samples. Since n represents the time delay as a number of samples, the value of n must be an integer. The corresponding incidence angle $\alpha = \frac{\pi}{2} - \gamma$ is calculated through Eq. (3.1). Table 5.3 shows the discernible angles for α which can be achieved with time delay values of 0 to 38. The table is symmetric for the negative time delay values.

n	α	e	n	α	e
0	0.0	0.75	20	31.4	0.88
1	1.5	0.75	21	33.2	0.90
2	3.0	0.75	22	35.0	0.92
3	4.5	0.75	23	36.9	0.95
4	6.0	0.75	24	38.7	0.97
5	7.5	0.75	25	40.7	1.00
6	9.0	0.76	26	42.7	1.03
7	10.5	0.76	27	44.8	1.07
8	12.0	0.77	28	46.8	1.12
9	13.6	0.77	29	49.1	1.17
10	15.1	0.78	30	51.5	1.23
11	16.7	0.78	31	53.9	1.31
12	18.2	0.79	32	56.6	1.41
13	19.8	0.80	33	59.4	1.54
14	21.4	0.81	34	62.5	1.71
15	23.0	0.82	35	65.9	1.98
16	24.7	0.83	36	69.8	2.46
17	26.3	0.84	37	74.8	3.76
18	28.0	0.85	38	82.3	3.76
19	29.7	0.87			

Table 5.3: Discernible angles for time delays in integer units for the air experimental setup. n: Time delay in integer units. α: Angle α in degrees. e: Maximum quantization error in degrees.

Table 5.3 shows the quantization error due to the discretization of time delays. The theoretical resolution of approximately 1.5^o or corresponding maximum quantization error of $\pm 0.75^o$ is accurate only at a time delay of 1. The error is close to that value when the time delay is within 30 samples (51.5^o). Beyond that the error starts to grow. Therefore,

in our setup, the error caused by the quantization dominates the far-field assumption error, since only when the distance from the source to the microphones is less than about $0.25m$, the far-field assumption error starts to be greater than quantization error.

It is interesting to observe the range of values of r/b (distance to baseline ratio) that give errors due to violation of the far-field assumption comparable to the quantization errors due to discretization of delay measurements. The range of values of r/b computed based on Table 5.3 and Eq. (3.10) is $[0.67, 1.75]$ for the air experiments. The worst case, corresponding to the highest r/b, is in the vicinity of 45^o, corresponding to the maxima in Figure 3.3.

Figure 5.9 shows the waveform of the audio data at the two channels and the output of the correlation function.

5.2.2 Results

In the air experiments, the sound source is placed at a fixed position and we estimate its position 10 consecutive times. Then the sound source is moved and the experiment is repeated at the new position. A total of 25 sound source positions throughout the area around the two microphone arrays are chosen and for each sound source position, mean and standard deviation are calculated as two measures to illustrate the consistency and the accuracy of our approach.

We use the same setup and process to test our method described in Chapter 3 and Rabinkin's method [13]. . A summary of Rabinkin's localization algorithm is described in section 2. The results of both approaches are shown in Figure 5.10. The cross is centered at the average position of those measurements. The height and width of each cross represent the standard deviation along the two dimensions. We observe that the standard deviation on z-axis is generally greater than x-axis and y-axis. This is also reported in [13] and explained by the fact that the baseline along the z dimension is smaller.

The original method [13] requires signals to be synchronously digitized from every microphone, both within and across microphone arrays. This would be very difficulty for arrays at some distance from each other. Another problem is that when more arrays are added to improve area coverage or accuracy of localization, ever more sophisticated centralized digitization hardware to accommodate additional channels, and ever longer cables to transmit the analog sound to the digitization hardware are required. In our method, we only compute time delays between microphones that are close to each other. Therefore, the possible time delay range is narrow and we can use small correlation windows to shut out reflections. In Rabinkin's method, which needs to estimate time delays between arrays, we have to enlarge the correlation window, making time delay estimates less robust. To address this problem,

we introduce a "decentralized" version of Rabinkin's method, by only estimating the time delays between the microphones within the same array. The standard deviation of the estimates is shown in Figure 5.10. We observe that there are certain source positions where "decentralized" Rabinkin's method fails to give a meaningful estimate [1] Those positions are represented by squares in Figure 5.10.

Figure 5.11 shows the means of the estimated positions and the actual sound source locations overlaid on top of each other, which gives us a graphic view of the distance between the estimates and the actual source positions. Table 5.4 is a summary of the average of the standard deviations, the average distance between the estimates and the actual positions, and standard deviation on the distances.

(unit: m)

	Our Approach			Rabinkin's Approach (Centralized)			Rabinkin's Approach (Decentralized)		
	x	y	z	x	y	z	x	y	z
δ	0.043	0.043	0.074	0.035	0.039	0.053	0.042	0.041	0.074
ϵ	0.041	0.040	0.084	0.044	0.045	0.063	0.050	0.049	0.077
$\delta(\epsilon)$	0.036	0.034	0.046	0.036	0.033	0.035	0.038	0.034	0.046

Table 5.4: This table shows a summary of experimental results of our approach, centralized and decentralized Rabinkin's approach, without the failed estimates. δ: Average of standard deviations. ϵ: Average error in estimates of source position. $\delta(\epsilon)$: Standard deviation of errors.

From our experimental results, we also observe, as in [13], that the estimates along the z dimension are less accurate than the other two dimensions. However, we did encounter convergence problems for some source locations with decentralized Rabinkin's method.

5.3 Pool Experiment

5.3.1 Experimental Setting

The underwater experiments were carried out in a standard swimming pool. The pool is divided into two parts by a wooden deck. We chose the deep end of the pool to test our approach, which is 25.0m long, 20.0m wide, and 4.3m deep at the flat bottom of the deep end as shown in Figure 5.12.

[1] The estimates of those source positions are over 100m, which exceeds the size of the room for the experiment.

The listening apparatus consists of four DolphinEar/PRO omnidirectional hydrophones, which are attached to the corners of a square buoy of size $1.0 \times 1.0m$, as shown in Figure 5.13.

The same digital recording system with the same parameter settings used in the air experiment is used for the audio data collection and computation in our pool experiment, as shown in Table 5.5.

Parameter	**Value**
Sampling Frequency	44100Hz
Sample Resolution	16 bits
Sample Size	2048
Low Pass Filter Cutoff Frequency	4000Hz
High Pass Filter Cutoff Frequency	200Hz
Filter Coefficients	128 (for low and high pass)
$\mathbf{V_{thresh}}$	Dynamically Determined

Table 5.5: Parameter Specifications for The Experiments in The Air

Two kinds of impulsive sound are used in the pool experiments. One is generated by a person floating at the surface, hitting a long copper pipe directly against the bottom of the pool. The other is by a diver lying on the bottom of the pool, who strikes two pieces of metal pipe against each other, as shown in Figure 5.14. The time difference between successive impulses is approximately $1.0s$. Figure 5.15 shows the spectrogram of the recorded sound. The frequency range of interest is $200 - 4000Hz$. Frequency not in this range will be filtered out.

In our pool experiment, the minimum distance from the hydrophone array to the sound source is $4.3m$. Therefore, the minimum ratio of the distance to the source and the length of hydrophone baseline is $r/b = 4.3$. The corresponding far-field error is shown in Figure 5.16 as a function of incidence angle α.

We use the method described in Sec. 5.2.1 to compute the range of the time delay for the sound signal in the water, using the speed of sound in the water $c = 1,482m/s$:

$$\pm n = \pm \frac{f|\mathbf{b}|}{c} \approx \pm 29 \tag{5.2}$$

Therefore, the estimate of the time delay from the correlation function of the pool experimental data must lie in the interval $[-29, 29]$. Any estimate not in this interval is considered to be a false peak and it is discarded. This interval is also used to narrow down the correlation window considered, which helps exclude reflections.

Table 5.6 shows the discernible angles for incidence angle α from Eq. (3.1) for all possible discrete time delay values. The time delay values are close to the theoretical resolution of approximately 1.9^o up to about time delay of 20 (42.2^o) and start to diverge from it after that number.

n	α	e	n	α	e
0	0.0	0.96	15	30.3	1.13
1	1.9	0.96	16	32.5	1.16
2	3.9	0.97	17	34.8	1.19
3	5.8	0.97	18	37.2	1.23
4	7.7	0.97	19	39.7	1.27
5	9.7	0.98	20	42.2	1.33
6	11.6	0.99	21	44.9	1.39
7	13.6	0.99	22	47.7	1.47
8	15.6	1.00	23	50.6	1.57
9	17.6	1.02	24	53.8	1.70
10	19.6	1.03	25	57.2	1.87
11	21.7	1.04	26	60.9	2.12
12	23.8	1.06	27	65.1	2.54
13	25.9	1.08	28	70.2	3.42
14	28.1	1.10	29	77.0	3.42

Table 5.6: Discernible incidence angles α in degrees for integer time delay values n for the water experiments. The maximum quantization error in degrees is e.

The range of values of r/b (distance to baseline ratio), that gives errors due to violation of the far-field assumption comparable to the quantization errors due to discretization of delay measurements, is [0.79, 1.66]. This is computed based on Table 5.6 and Eq. (3.10).

5.3.2 Results

In the pool experiments, the sound source is at a fixed position on the bottom with coordinates [11.59, 7.91, -4.30] (in m), as shown in Figure 5.12. We use only one hydrophone array to simulate multiple arrays as follows. First, we collect the audio data with the hydrophone array at one position on the surface, we estimate a number of time delays, and calculate a robust average (with outliers excluded). Then the hydrophone array is moved to a new position and we repeat the same process, thus simulating the availability of multiple hydrophone arrays. A sample of the intermediate results for data processing of the pool experiments is presented in Appendix C.

Figure 5.17 shows eight positions of the hydrophone array and the fixed position of

the sound source, for which data was collected. For each hydrophone array position, we construct a direction line of sound propagation towards the sound source. Assuming we have a stationary source, we combine multiple hydrophone array positions, with one direction line estimated at each position, thereby simulating different number and positions of hydrophone arrays.

The estimates of the z coordinate of the source position, shown in Appendix D, are much poorer than the estimates of the x and y coordinate. There is a systematic error present that causes the z estimates to be much closer to 0 than the actual z. This means that there are propagation paths not accounted for in our simple direct-path model of propagation.

If we assume the availability of a depth sensor, we can estimate the sound source position to within 0.8m by averaging estimates from several hyddrophone array positions. Table 5.7 shows the intersections of the direction vectors estimated at the eight positions of the hydrophone array with the horizontal plane where the actual source position lies on ($z = -4.30m$). The corresponding plot of those intersections is shown in Figure 5.18. The differences between the actual source position and the average estimate of the intersections on the plane $z = -4.30m$ with the two kinds of sound are $\Delta x = 0.17m, \Delta y = 0.60m$ and $\Delta x = 0.80m, \Delta y = 0.10m$ respectively.

(unit: m)

Hitting-long			Hitting-short		
	x	y		x	y
G1	11.23	3.03	G1	10.07	8.07
G2	14.27	12.66	G2	10.63	8.10
G3	12.86	10.86	G3	10.50	8.80
G4	9.48	9.51	G4	10.04	9.94
G5	11.55	9.62	G5	12.53	8.13
G6	11.20	7.35	G6	11.48	7.78
G7	9.93	5.10	G7	11.04	6.15
G8	10.82	9.92	G8	10.05	5.49
mean	11.42	8.51		10.79	7.81
σ	1.54	3.16		0.87	1.41
mean absolute error	0.17	0.60		0.80	0.10

Table 5.7: This table shows the intersections of the direction lines estimated at the eight hydrophone array positions (shown in Figure 5.17) with the horizontal plane ($z = -4.30m$) in the pool experiment. The actual source position is [11.59, 7.91, -4.30]. The left three columns are the results from using a long pipe striking on the bottom, and the right three are the results from striking two short pieces of metal pipe against each other at the bottom.

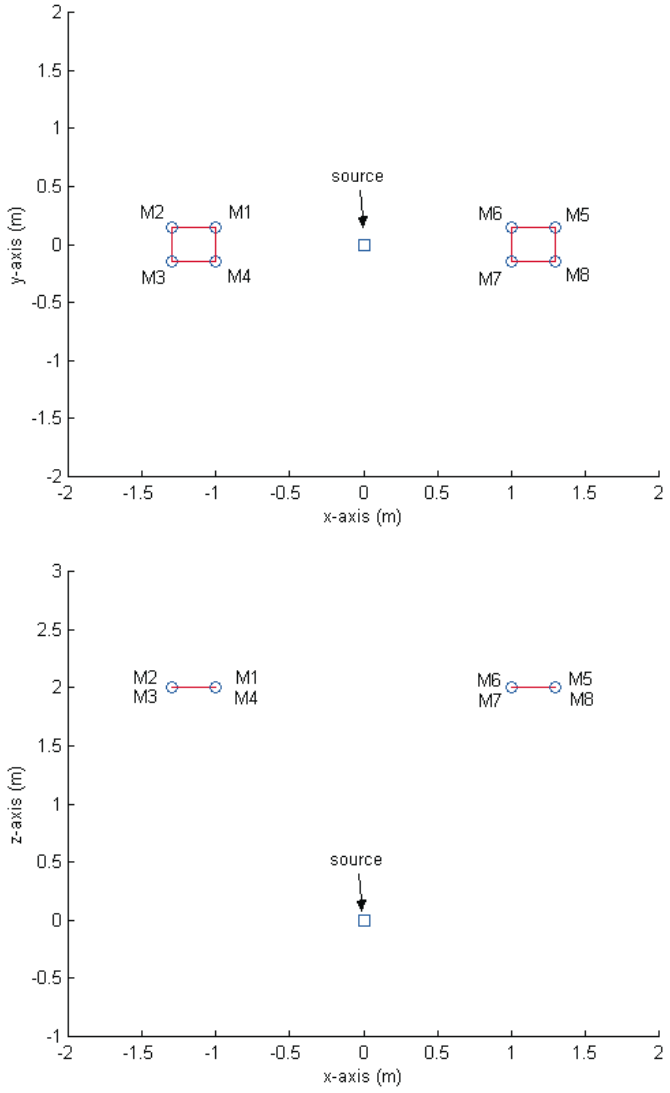

Figure 5.2: Microphone arrangement in the simulation program. Eight microphones are represented as circles and the source is represented as a square.

```
Algorithm: Random error in time delay estimate change algorithm
Input: Microphone positions, source position, maximum range of time
    delay estimation error in samples, number of measurements.
Output: Mean of estimates, standard deviation.

Method:
(1):  initialize the range of error to be 0;
(2):  repeat:
(3):      for each measurement
(4):          compute the time delays based on the microphone
                  and the actual source positions;
(5):          for each time delay estimate
(6):              generate a random error in the specified range;
(7):              add the error to the time delay estimates;
(8):          estimate the source position;
(9):      compute mean and standard deviation;
(10):     increase the range of error by one sample;
(11): until the range of error exceeds the maximum range.
```

Figure 5.3: Random error in time delay estimation change algorithm. The term *random* refers to a function which would produce a uniformly distributed random number between $[-range, range]$.

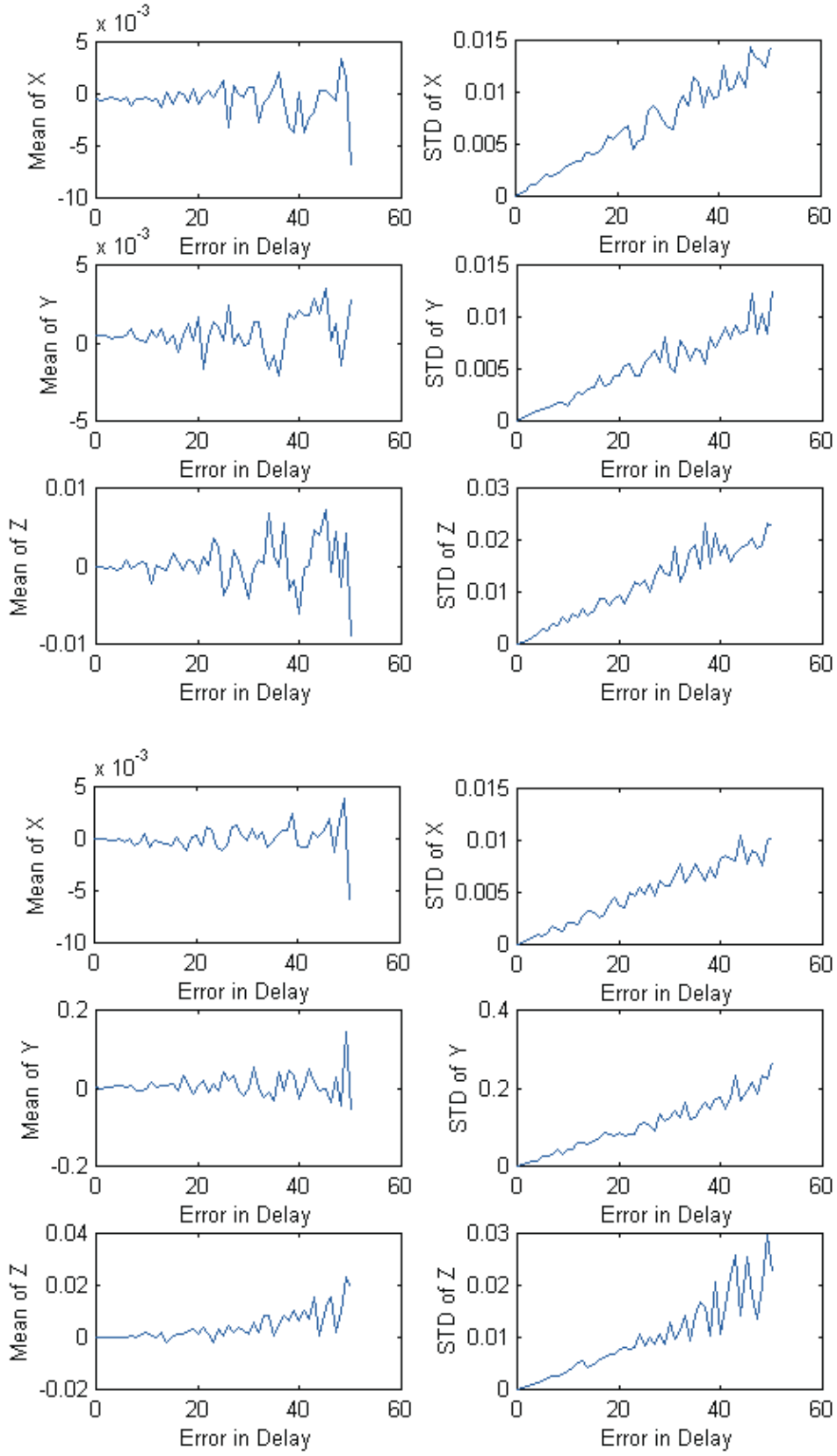

Figure 5.4: The plots illustrate the performance of our approach and Rabinkin's approach [13] in the simulation program, where a random time estimate error is added. The top six graphs are the average and the standard deviations of the estimates of our approach, and the bottom six are the results of Rabinkin's approach. Note the different vertical scales.

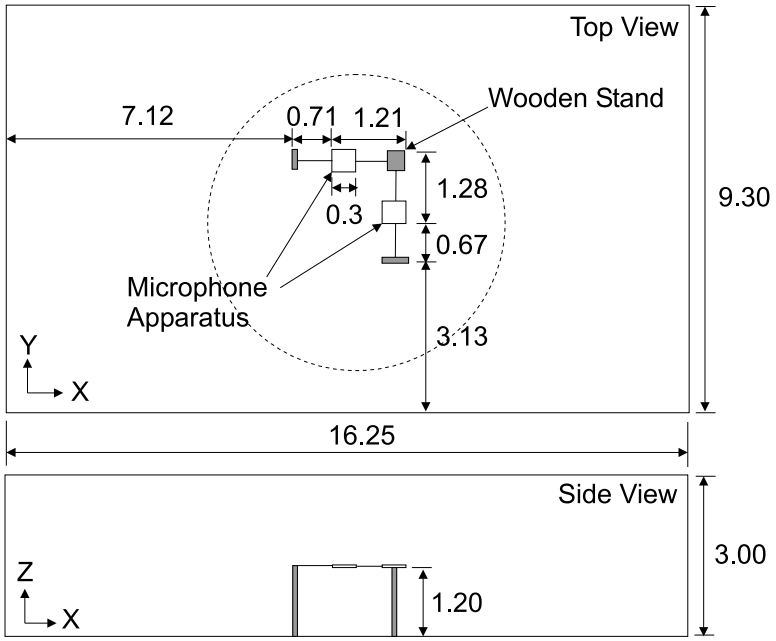

Figure 5.5: Room arrangement for the air experiments. The size of the room is $16.25 \times 9.30 \times 3.00m$. Two microphone arrays are placed at a height of $1.2m$ from the ground. The interior of dotted area represents possible positions of the sound source, which is at a height of $0.02m$ above ground.

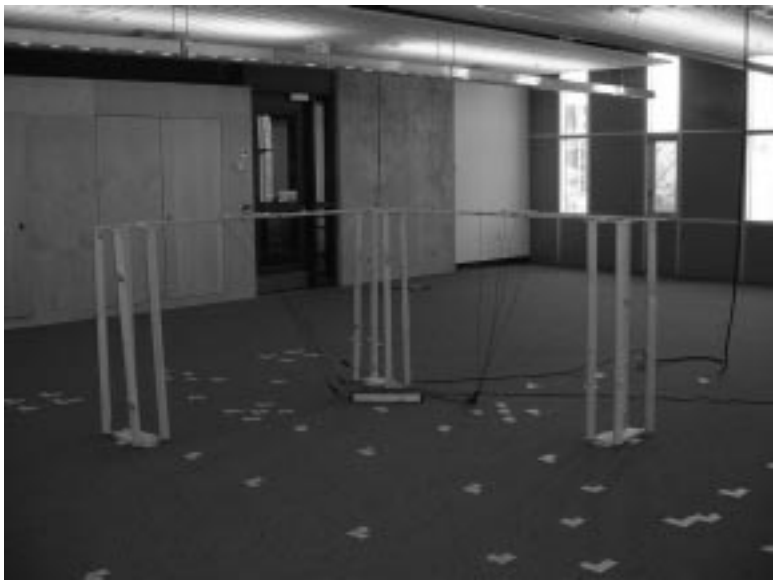

Figure 5.6: Actual photograph of the room arrangement where the air experiments took place.

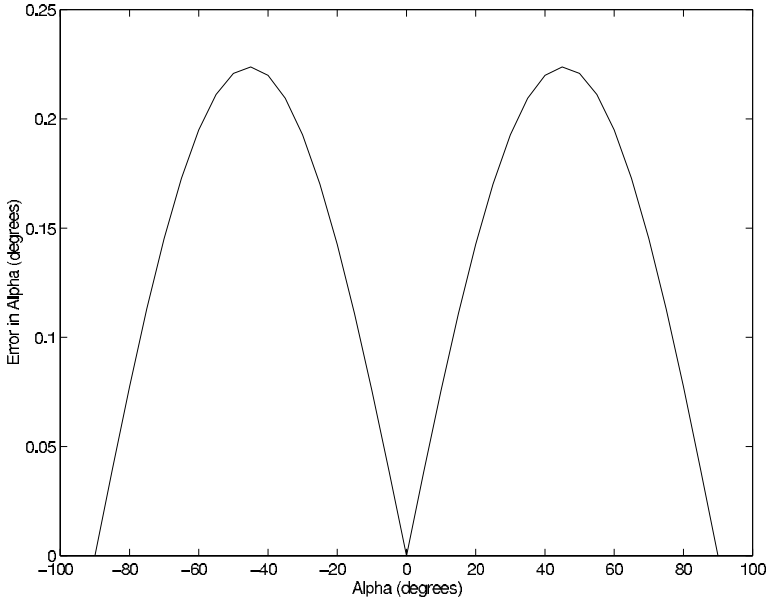

Figure 5.7: This figure shows the far-field assumption error at the value of $r/b = 4$, which is used as minimum value of r/b in our air experiments. It represents the error introduced by making far-field assumption.

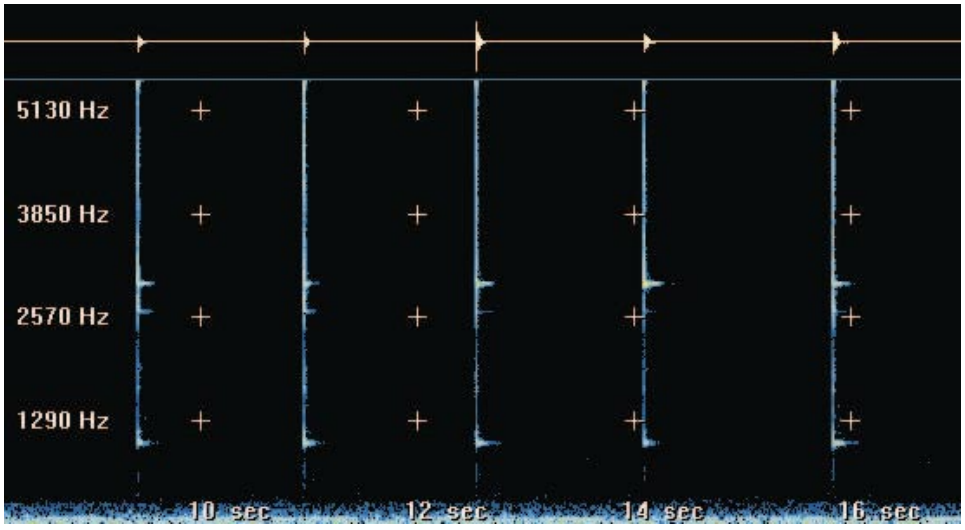

Figure 5.8: The spectrogram of the impulsive sound in the air experiments. Background noise is at frequency below $200Hz$. Frequency not in the range of $200 - 4000Hz$ will be filtered out by the bandpass filtering.

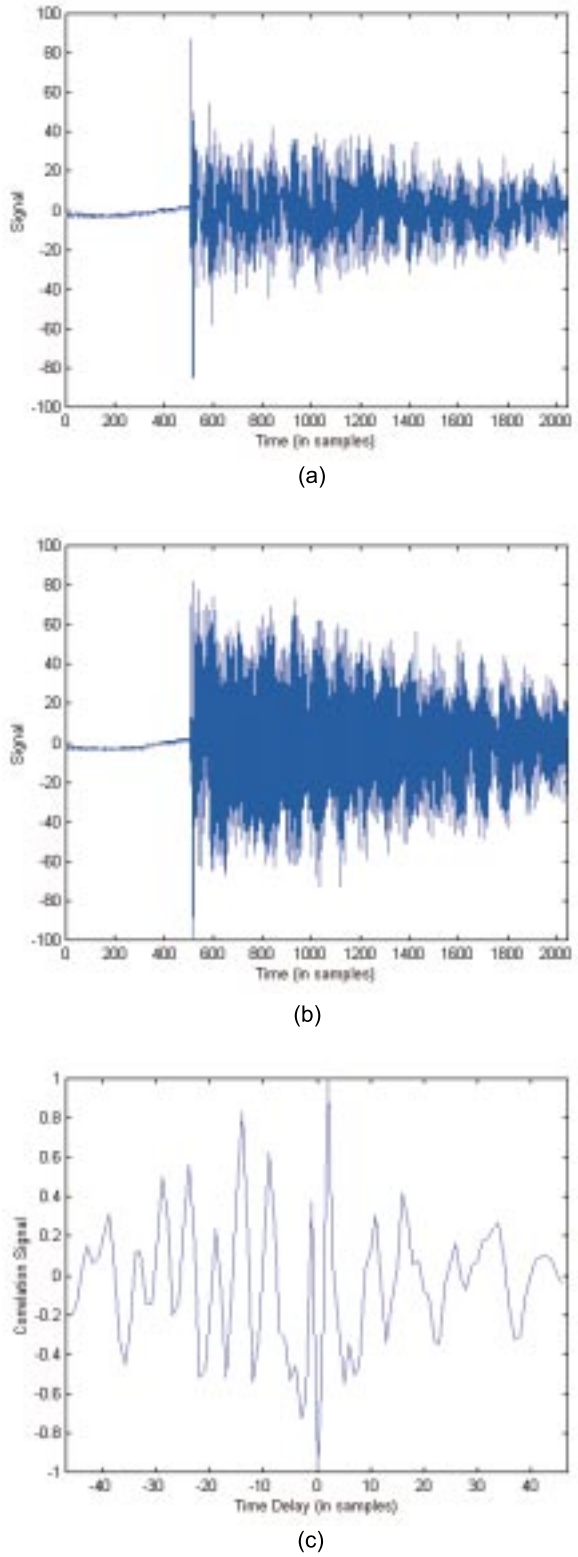

Figure 5.9: The plots show the waveform of data from two channels in the air experiment and their correlation. The sound recorded is of a single strike on a metal appliance. The strongest peak in the correlation function represents the time delay (2 samples). (a): Waveform of data from first channel. (b): Waveform of data from second channel. (c): Output of correlation function.

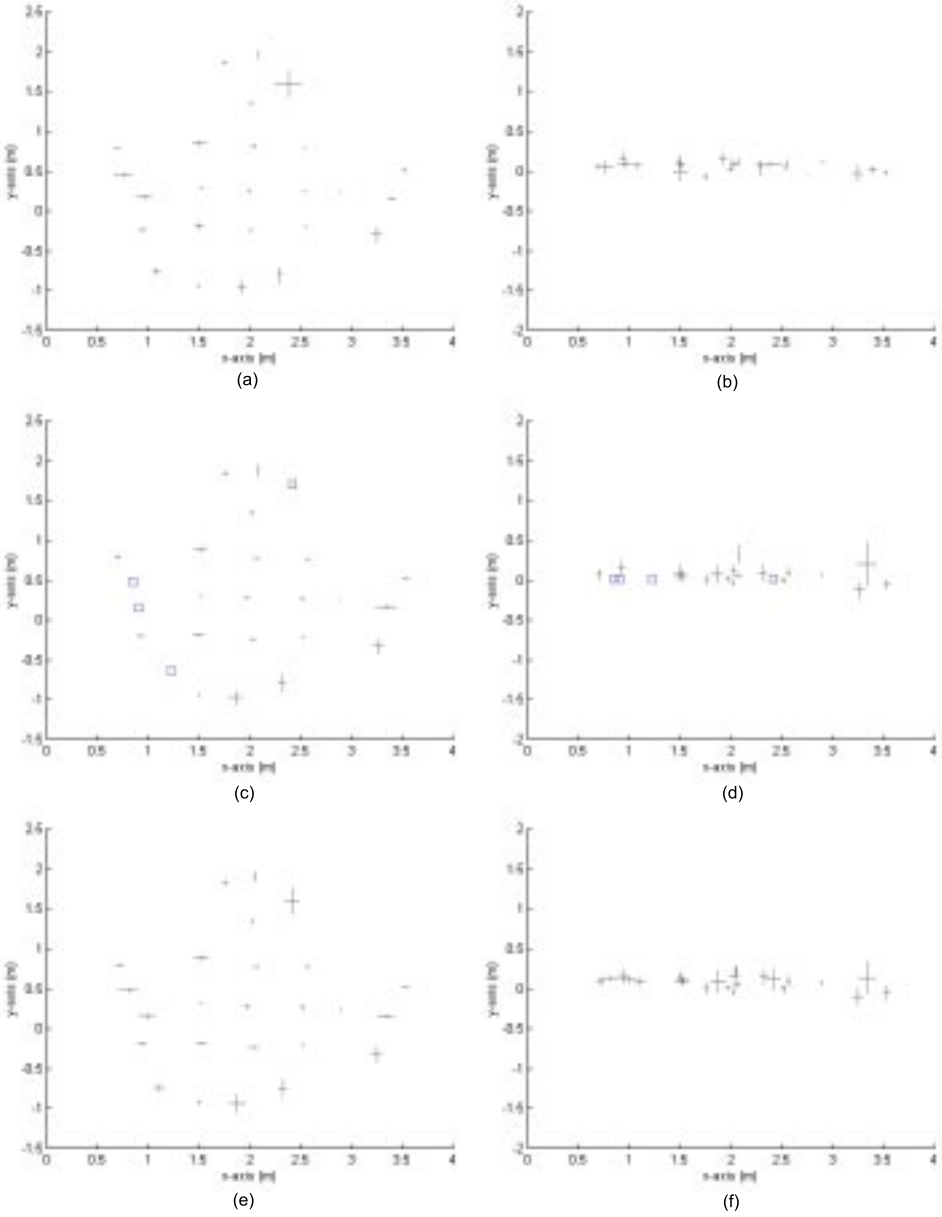

Figure 5.10: The graphs show the estimated sound source position of the air experiments in x-y plane and x-z plane, corresponding to the layout of the room as shown in Figure 5.5. The size of the cross represents the standard deviation. The center of the cross is the estimated mean position of the sound source. (a) and (b): Centralized Rabinkin's method [13]. (c) and (d): Decentralized Rabinkin's method. (e) and (f): Standard deviation on x,y and x,z axes of our approach.

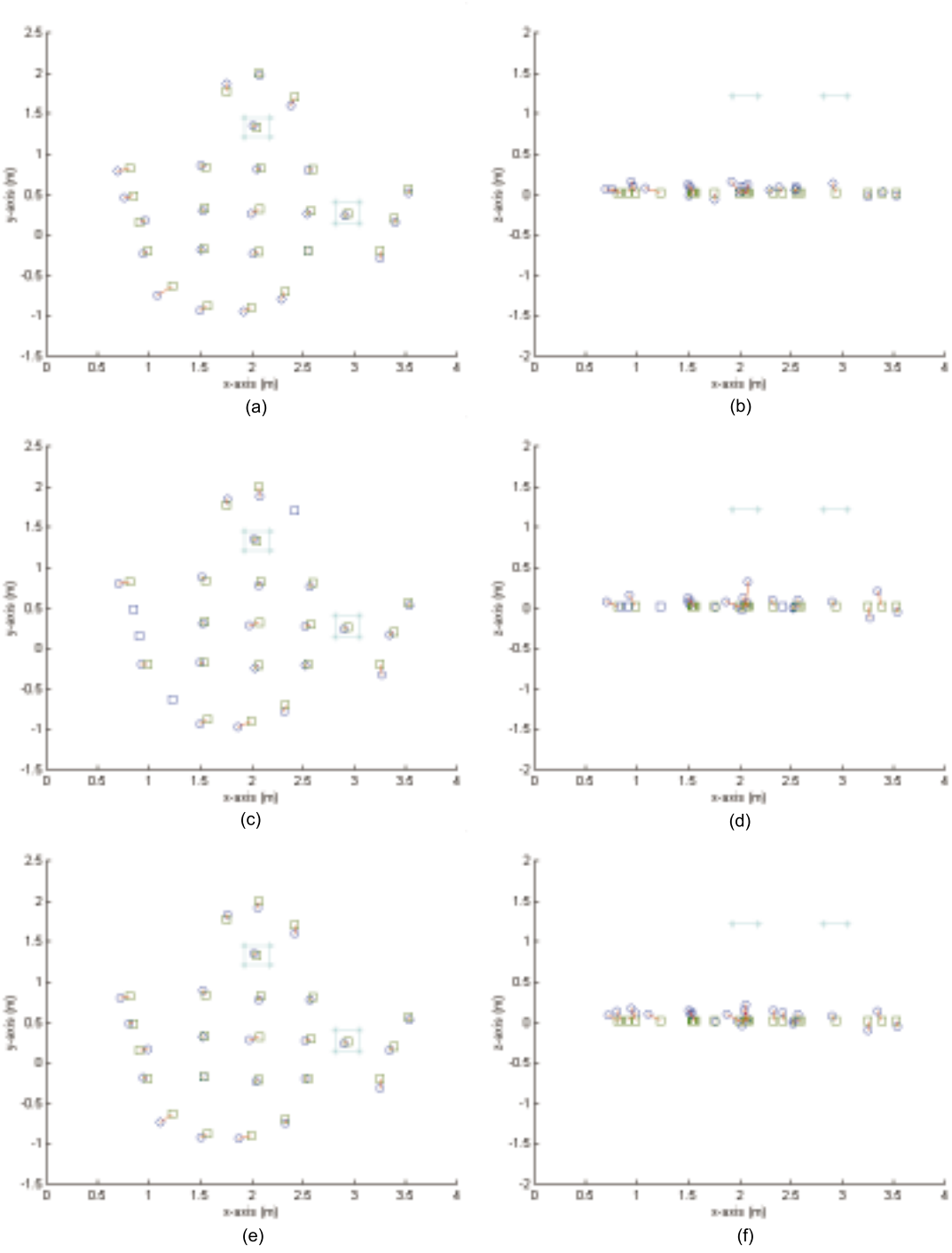

Figure 5.11: The graphs overlay the actual source positions and the estimated mean positions from the air experiments, along the x,y axes and x,z axes (a) and (b): Centralized Rabinkin's approach [13]. (c) and (d): Decentralized Rabinkin's approach. (e) and (f): Our approach

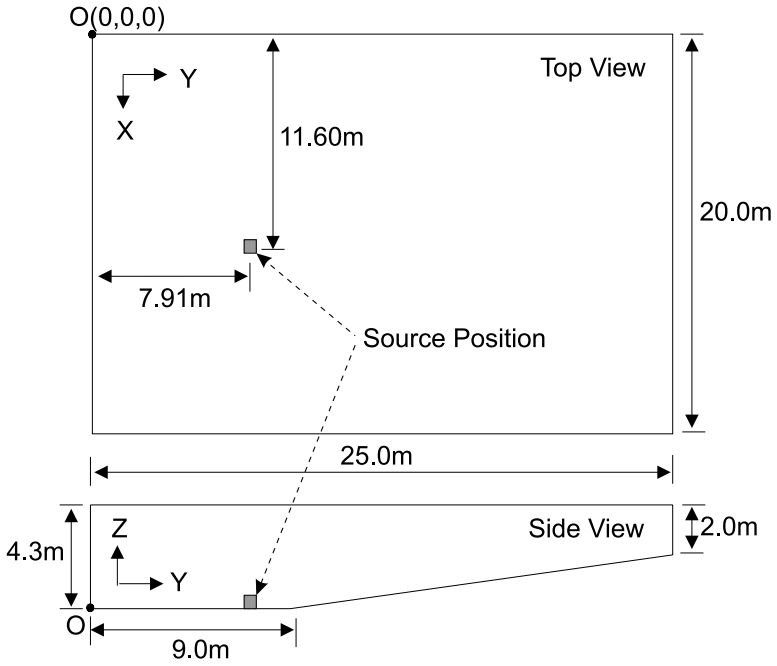

Figure 5.12: Profile of the deep end of the pool for the experiment in the water. The sound source is placed at the flat bottom of the pool.

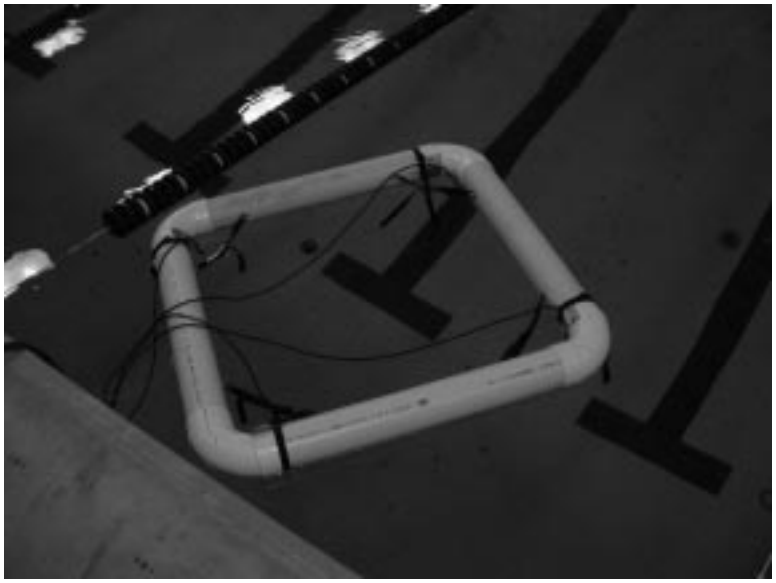

Figure 5.13: Four omnidirectional hydrophones attached on a square buoy ($1.0 \times 1.0m$).

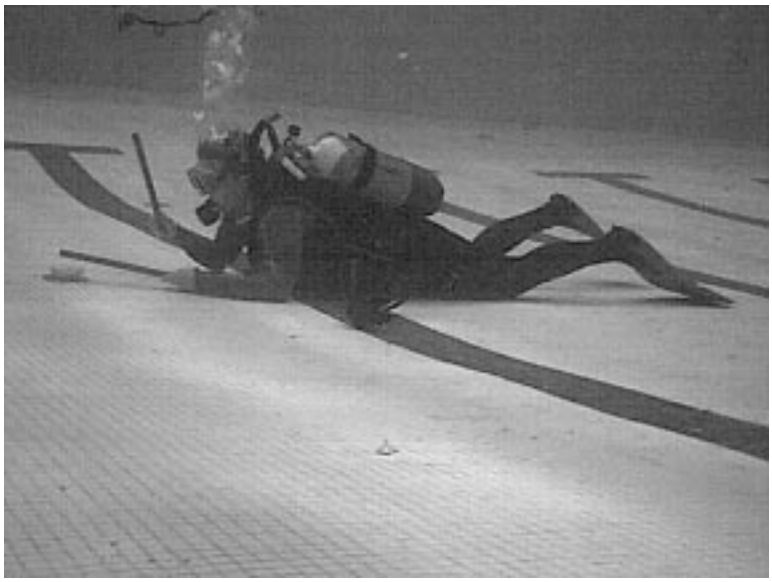

Figure 5.14: The sound is generated by a person who dives into the bottom of the pool, striking two pieces of metal pipes against each other, at intervals of approximately 1.0s.

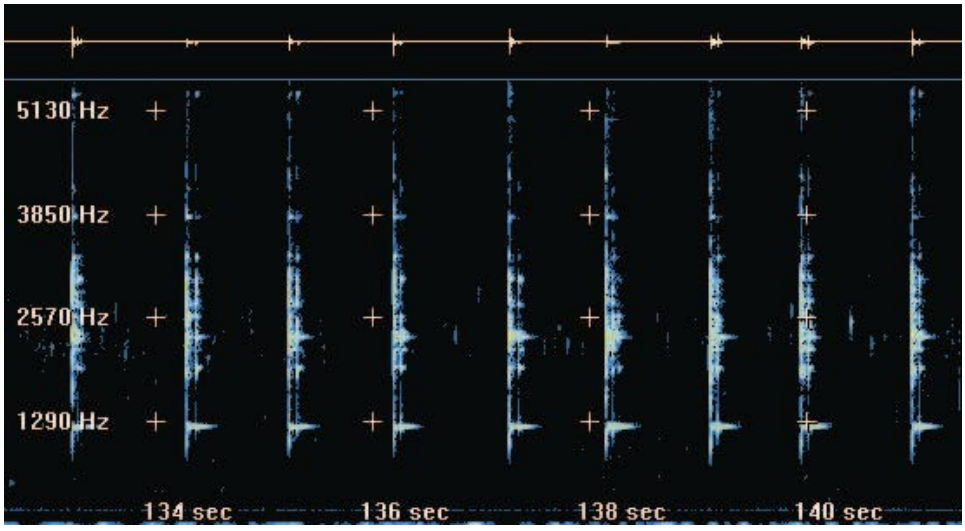

Figure 5.15: The spectrogram of an impulsive sound in the pool experiment, generated by a diver on the bottom striking two pieces of metal pipes against each other. Frequency not in the range of $200 - 4000Hz$ will be filtered out by the bandpass filtering.

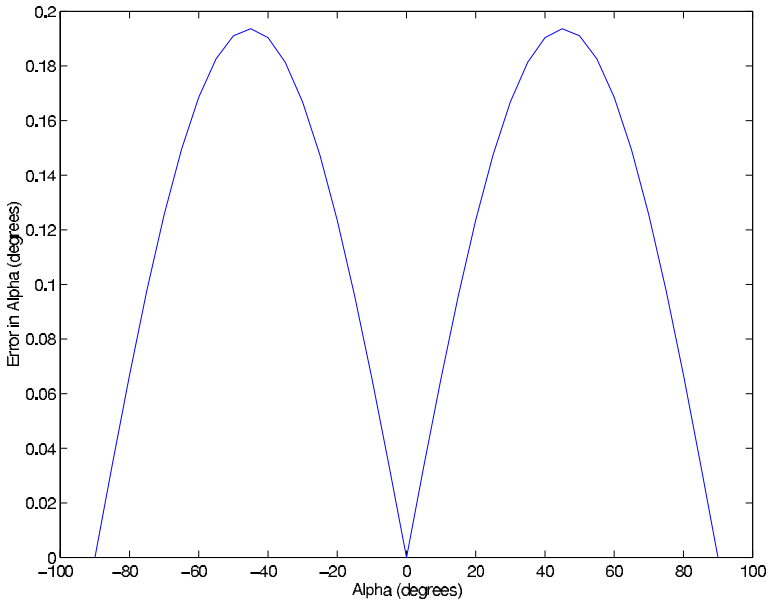

Figure 5.16: This figure shows the error due to the far-field assumption at the minimum value of $r/b = 4.3$ as a function of incidence angle α.

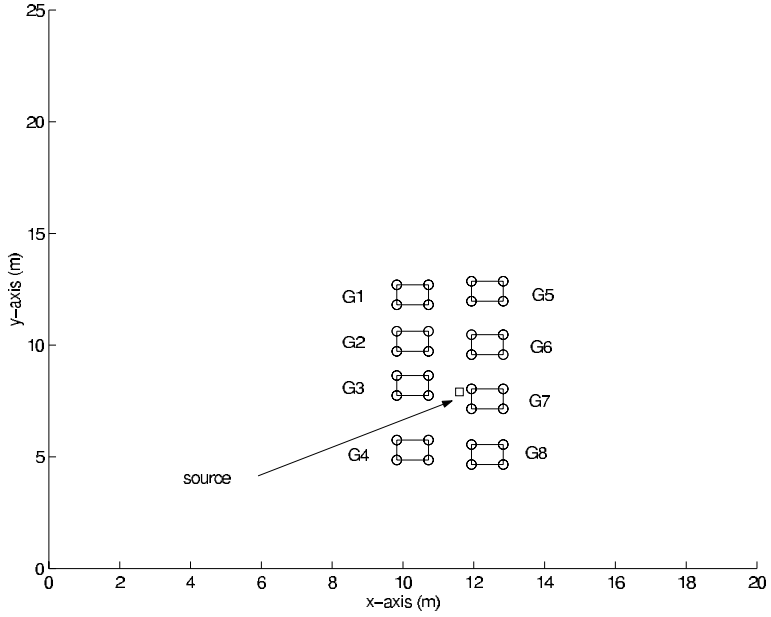

Figure 5.17: The diagram shows the eight hydrophone array locations ($G1 - G8$) used in the pool experiment on the x, y plane, where y is along the length of the pool starting from the deep end and x is along the width of the pool. The z axis is perpendicular to the x, y plane and directed upwards. The hydrophones are at the surface of the water with $z = 0$, and the source is at the bottom with $z = -4.3m$.

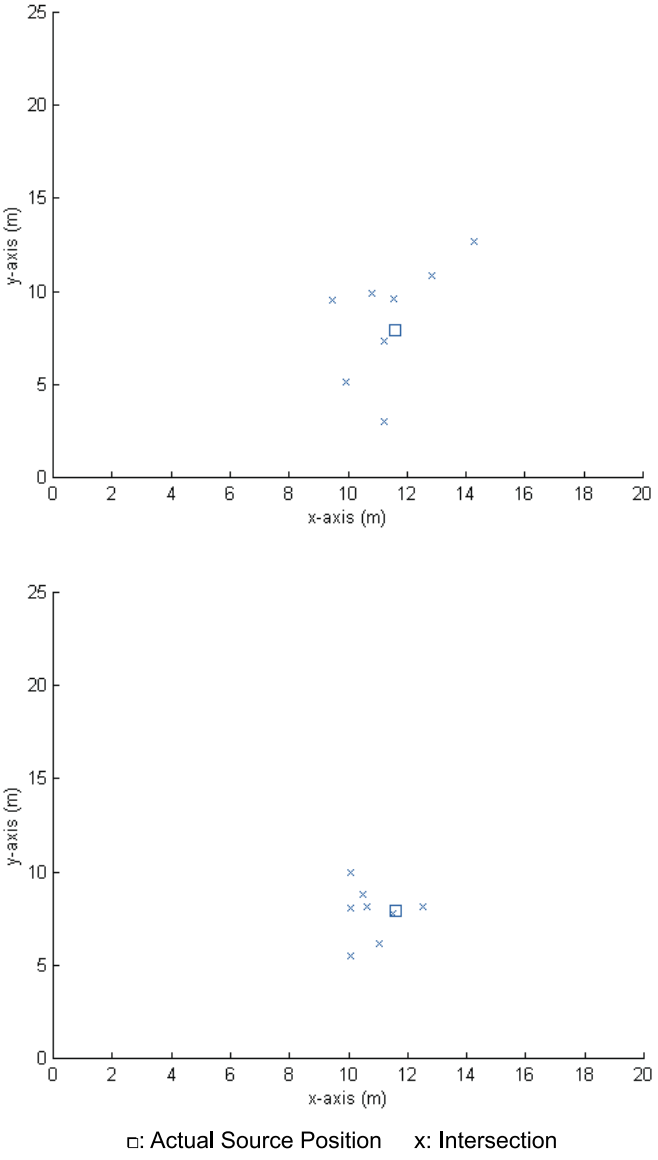

Figure 5.18: This plot shows the intersections of the direction lines estimated at the eight positions (shown in Figure 5.17) with the horizontal plane ($z = -4.30m$) in the pool experiment. The top figure is the result from using a long pipe striking on the bottom, and the bottom figure is the result from striking two short pieces of metal pipe against each other at the bottom.

Chapter 6

Discussion

In this report, we have presented a technique that uses the far-field assumption to estimate the direction of arrival of sound at a microphone array. We tested the method both in the air, using microphone arrays, and in the water, using a hydrophone array floating at the surface. The optimal location of the sound source is estimated as the intersection of the direction lines from each array pointing toward the source. Our technique operates in the audio frequency range, and therefore it can be implemented using off-the-shelf audio hardware. Compared with other systems operating in the ultrasonic range, signal attenuation is lower and audio hardware is low-cost, at the expense of decreased spatial resolution.

The motivation of our research is the design of an acoustic localization system for an underwater walking robot. A sound source will be mounted on the robot, and the robot position will be estimated with respect to the hydrophone arrays at the surface. The absolute positions of the hydrophone arrays can be determined by GPS. Their relative position can be determined more accurately by optical methods, such as those described in [15].

We tested our technique both in the water, and in the air, with a set up that reflects the distances of the water experiment scaled for the difference in the speed of sound. The air experiment allows us to confirm the validity of the method, in a more manageable experimental setting compared to the water. The results of both the air and the pool experiments are reported. The average accuracy in the air experiment is $0.05m$. In order to achieve reasonably accurate results in the water experiments, an additional depth sensor is required, that gives an estimate of the depth of the source, resulting in position accuracy of $0.8m$.

Rabinkin's approach described in [13] was implemented and tested with the same setup of our air experiment, and the result is comparable to that of our method. But it requires the synchronization and digitization of signals between microphones in different arrays, which needs centralized digitization hardware. When we tried to decentralize Rabinkin's method, the method failed to converge in the localization of some source positions.

We further discuss two more geometric models in the appendix. One uses a hydrophone pair to estimate the angle of arrival of sound, which corresponds to a cone in 3D space. Sound source position is estimated as the intersection of multiple cones constructed by multiple hydrophone pairs. Yet another localization approach discussed uses a combination of a hydrophone array and hydrophone pair. Simulation results from these methods are presented.

Future work includes the testing of the algorithms presented in the appendix on real data; the testing of all our acoustic positioning methods in an open water environment; the estimation of time delays with sub-sample accuracy by fitting a quadratic to the correlation function. Further investigation of the systematic errors in the z estimate in the water experiments is required. Signals at hydrophones near the air-water boundary are the superposition of the direct path signal and its reflection off the air-water boundary, which is subject to a phase shift of π. This observation suggests placing the hydrophones at mid-depth instead, an experiment that requires a larger and deeper water body than the pool. Our method may still be useful for estimating the x, y coordinates of the source in shallow water, where waveguide effects are expected between the bottom and the air-water boundary.

To complete our underwater robot localization system, we need to design an underwater sound source. Our localization method needs to be integrated into a data fusion system for robot positioning, including depth sensors, inertial sensors, 3DOF compass, visual positioning, and dynamic modelling of the robot motion.

Finally, we are interested in mounting the hydrophone arrays onto autonomous self-contained motorized buoys, capable of navigation, and direction of arrival estimation, which they will communicate to each other using wireless ethernet links. Availability of such robotic buoys would enable the positioning of the hydrophone arrays with respect to the underwater robot so as to optimize localization accuracy.

Bibliography

[1] P. Gerstoft, W. Hodgkiss, W. Kuperman, H. Song, M. Siderius, and P. Nielsen. Adaptive beamforming of a towed array during a turn. *IEEE Journal of Oceanic Engineering*, 28(1):44–54, Jan. 2003.

[2] D.A. Gray, B.D.O. Anderson, and R.R. Bitmead. Towed array shape estimation using kalman filters-theoretical models. *Oceanic Engineering, IEEE Journal of*, 18:543–556, October 1993.

[3] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 1 edition, 2001.

[4] CMT Inc. *Introduction to the Global Positioning System for GIS and TRAVERSE*. http://www.cmtinc.com/gpsbook/index.htm, last accessed on July 27, 2003.

[5] B. Kapralos, M. Jenkin, E. Milios, and J. Tsotsos. Eyes' n ears: face detection utlizing audio and video cues. *Second International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems (RATFG-RTS 2001). In conjunction with ICCV 2001.*, 2001.

[6] A. P. Klimley, B. J. Le Boeuf, K. M. Cantara, J. E. Richert, S. F. Davis, and S. V. Sommeran. Radio-acoustic positioning as a tool for studying site-specific behavior of the white shark and other large marine species. *Marine Biology*, 138(2):429–446, 2001.

[7] Cho-Chung Liang, Tso-Liang Teng, and Wen-Hui Cheng. A study of a short-baseline acoustic positioning system for offshore vessels. *Marine Geodesy*, 22(1):19–30, 1999.

[8] B.T. Luke. *Agglomerative Clustering*. http://fconyx.ncifcrf.gov/ lukeb/agclust.html, last accessed on July 8, 2003.

[9] D.L. McKeown. Acoustic positioning of oceanographic instruments. *Ocean'74*, 2, August 1974.

[10] NOAA's (National Oceanic, Atmospheric Administration)Pacific Marine Environmental Laboratory, and National Marine Mammal Laboratory. Whale acoustics project publications, 2003. Newport, Oregon and Seattle, Washington.

[11] Jan Opderbecke. At-sea calibration of a USBL underwater vehicle positioning system. *Proceedings of Oceans'97*, 1:721–726, 1997.

[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Press Syndicate of the University of Cambridge, 2 edition, 1992.

[13] D. V. Rabinkin. *Optimum Sensor Placement for Microphone Arrays*. PhD thesis, Rutgers University, May 1998.

[14] G. L. Reid and E. Milios. Active stereo sound localization. *The Journal of the Acoustical Society of America*, 113:185–193, January 2003.

[15] Ioannis Rekleitis, Gregory Dudek, and Evangelos Milios. Experiments in free-space triangulation using cooperative localization. In *IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, pages 1777–1782, 2003.

[16] Defense Research and Development Canada (DRDC). Workshop on detection and localization of marine mammals using passive acoustics, 19 - 21 November 2003.

[17] Rodney Rountree, Cliff Goudey, and Tony Hawkins, editors. *Listening to Fish: Proceedings of the International Workshop on the Applications of Passive Acoustics to Fisheries*. MIT Sea Grant, 2002.

[18] A.P. Scherbatyuk and Y.V. Vaulin. Integrated positioning system for autonomous underwater vehicle MT-88. *Oceans 94 OSATES*, 3:III.384–III.388, September 1994.

[19] S.M. Smith and D. Kronen. Experimental results of an inexpensive short baseline acoustic positioning. *Proceedings of Oceans'97*, 2:V11.1–V11.6, October 1997.

Appendix A

Other Approaches for Sound Source Localization

A.1 Sound Source Localization Using Hydrophone Pair

A different method to localize the sound source is to use multiple hydrophone pairs, with each estimating an angle of arrival of sound direction, which corresponds a cone in 3D space. The sound source location is estimated as the intersection of multiple cones. Consider a hydrophone pair with two hydrophones. The angle of arrival of sound γ at a hydrophone pair can be estimated by Eq. 3.1. This angle of arrival constrains the position of the sound source to be on the surface of a cone. The apex of the cone is the center of the baseline between the two hydrophones, and its axis is the baseline itself. Similarly to what we describe in Sec. 3, the least mean squares criterion is used to solve for the optimal intersection in the presence of measurement noise. The optimal intersection is the point that has the total overall squared distance from the multiple cones.

A cone-centered coordinate system is defined as having its origin $\mathbf{O'}$ at the apex of the cone and as its x-axis the axis of the cone (Fig. A.1). The unknown source location $\mathbf{P}$ has coordinates $= [x', y', z']$ in this system.

Consider next a plane defined by point $\mathbf{P}$ and the x-axis. The intersection of this plane and the cone consists of two lines l_1 and l_2, passing through the origin and along the directions of unit vectors $\mathbf{a_1} = [a_{1x}, a_{1y}, a_{1z}]$ and $\mathbf{a_2} = [a_{1x}, a_{1y}, a_{1z}]$ (Fig. A.2). Then the computation of distance d of the point $\mathbf{P}$ to the cone becomes the computation of the shorter distance of $\mathbf{P}$ to the two lines, which is expressed as follows:

$$d = \min\{d_1, d_2\} \tag{A.1}$$

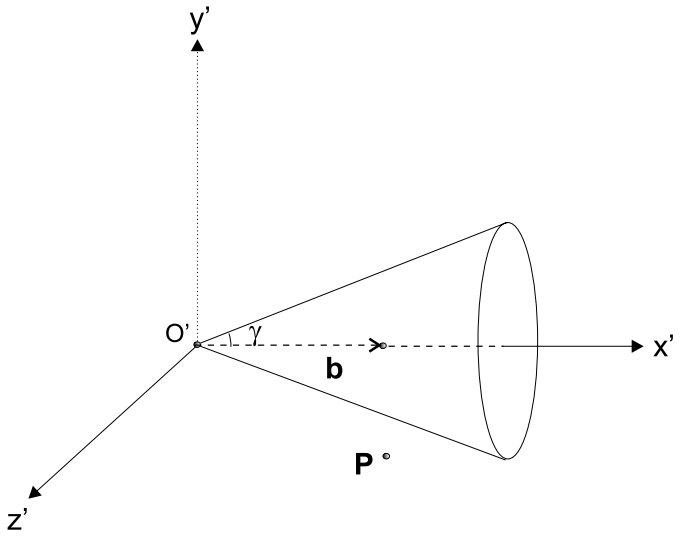

Figure A.1: The angle of arrival γ defines a cone-centered coordinate system, where the origin of the system is the apex of the cone, and the x-axis is the cone axis.

where d_1 and d_2 are the distances of $\mathbf{P}$ to l_1 and l_2 respectively, representing the error in the estimate of $\mathbf{P}$, as the lines should ideally contain $\mathbf{P}$.

From the 2D perspective, $\mathbf{a_1}$ and $\mathbf{a_2}$ actually are the baseline unit vector $\mathbf{b} = [1, 0]$ rotated by angle γ clockwise and counterclockwise, as shown in Fig. A.2. They can be calculated by applying the following rotation matrix:

$$\begin{cases} \mathbf{a_1} = \mathbf{R}(\gamma) \cdot \mathbf{b} \Rightarrow \begin{bmatrix} a_{1x} \\ a_{1y} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \\ \mathbf{a_2} = \mathbf{R}(-\gamma) \cdot \mathbf{b} \Rightarrow \begin{bmatrix} a_{2x} \\ a_{2y} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \end{cases} \tag{A.2}$$

or equivalently,

$$\begin{cases} \mathbf{a_1} = [\cos\gamma, \sin\gamma] \\ \mathbf{a_2} = [\cos\gamma, -\sin\gamma] \end{cases} \tag{A.3}$$

Therefore, the coordinates of $\mathbf{P}$ become $[x', \sqrt{(y'^2 + z'^2)}]$ on the above plane.

Consider the distance d_1 between $\mathbf{P}$ and line l_1. Point $\mathbf{P_0}$ is the projection of $\mathbf{P}$ onto line

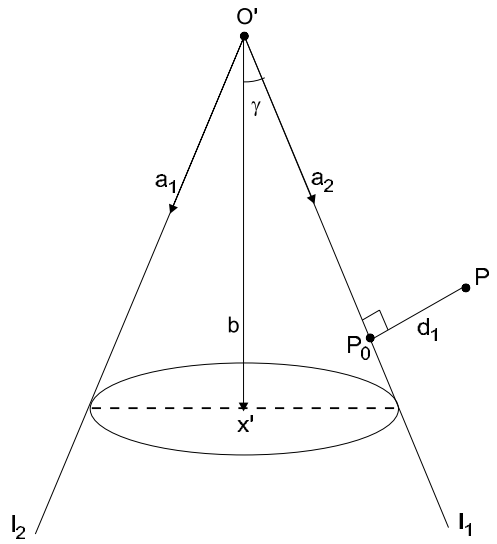

Figure A.2: This figure shows a plane defined by the point $\mathbf{P}$ and the x-axis. The intersection lines of the plane and the cone are l_1 and l_2, with starts at the origin $\mathbf{O'}$ and along the direction of $\mathbf{a_1} = [a_{1x}, a_{1y}, a_{1z}]$ and $\mathbf{a_2} = [a_{2x}, a_{2y}, a_{2z}]$. d_1 is the distance between $\mathbf{P}$ and l_1.

l_1, as shown in Figure A.2. Therefore, $\mathbf{P}$, $\mathbf{P_0}$ and $\mathbf{a_1}$ satisfy the constraint:

$$(\mathbf{P} - \mathbf{P_0}) \cdot \mathbf{a_1} = 0 \tag{A.4}$$

And point $\mathbf{P_0}$ lying on the line l_1 satisfies the following equation for some real number λ:

$$\mathbf{P_0} = \lambda \mathbf{a_1} \tag{A.5}$$

If we combine Eq. A.4 and Eq. A.5 and eliminate $\mathbf{P_0}$, then λ is expressed in terms of $\mathbf{P}$ and $\mathbf{a_1}$:

$$\lambda = \mathbf{P} \cdot \mathbf{a_1} \tag{A.6}$$

or equivalently,

$$\lambda = x' \cos\gamma + \sqrt{(y'^2 + z'^2)} \sin\gamma \tag{A.7}$$

Therefore, the distance square d_1^2 can be obtained by expression:

$$d_1^2 = d^2(\mathbf{O} - \mathbf{P}) - \lambda^2 \tag{A.8}$$

or equivalently,

$$d_1^2 = x'^2 \sin^2\gamma + (y'^2 + z'^2)\cos^2\gamma - 2x'\sqrt{y'^2 + z'^2}\sin\gamma\cos\gamma \tag{A.9}$$

Similarly, the distance square d_2^2 between point $\mathbf{P}$ and line l_2 can be obtained by:

$$d_2^2 = x'^2 \sin^2\gamma + (y'^2 + z'^2)\cos^2\gamma + 2x'\sqrt{y'^2 + z'^2}\sin\gamma\cos\gamma \tag{A.10}$$

The next step is to transform the current cone-centered system to the absolute coordinate system with a translation and rotation matrix. If we make the assumption that the z-axis is perpendicular to the plane of water surface in the absolute system, and the hydrophones are placed at the surface, then the transformation matrix is in the following form with the baseline unit vector $\mathbf{b} = [b_x, b_y, 0]$, and coordinates of the origin $\mathbf{O'}$ of the cone-centered system $= [o_x, o_y, 0]$ in the absolute coordinate system:

$$\mathbf{M} = \begin{bmatrix} b_x & b_y & 0 & -b_x o_x - b_y o_y \\ -b_y & b_x & 0 & -b_x o_y + b_y o_x \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A.11}$$

The coordinates of the point $\mathbf{P}$ in the cone-centered system are related to the coordinates $[x, y, z]$ in the absolute system through Eq. A.11, yielding:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} b_x & b_y & 0 & -b_x o_x - b_y o_y \\ -b_y & b_x & 0 & -b_x o_y + b_y o_x \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{A.12}$$

The function of the distance square d^2 from $\mathbf{P}$ to a single cone with respect to the three unknown variables x, y and z is obtained from Eq. A.12.

Therefore, with given multiple hydrophone pairs, corresponding to multiple cones, the sound source position can be determined by solving the following minimization problem:

$$\min_{x,y,z} \Big(\sum_i d_i^2 \Big) \tag{A.13}$$

where d_i is the distance from the source position to one cone determined by a hydrophone pair. Numerical methods can be used to solve this minimization problem [12].

A.2 Sound Source Localization Using Combination of Hydrophone Pair and Hydrophone Array

In this method, hydrophones are grouped into pairs and arrays, and the problem of locating the sound source position is defined as finding an optimal point which has minimum total squared distance to the cones and direction lines constructed by the hydrophone pairs and arrays in 3D space:

$$\min_{x,y,z} \Big(\sum_i d_i^2 + \sum_j r_j^2 \Big) \tag{A.14}$$

where d_i^2 is the distance squared from the source position to the i-th direction line, and r_j^2 is the distance squared to the j-th cone. They can be obtained by the methods described in Sec. 3 and Sec. A.1. This minimization problem can be accomplished through numerical means [12].

A.3 Simulation Results

A simulation program running on Matlab6.0 was carried out to test our positioning methods described in Appendix A.1 and Appendix A.2. The setup of the microphones is shown in Table A.1. Setup1 consists of two microphone arrays and one microphone pair and Setup2 consists four microphone pairs. The sound source is placed at the depth of $100m$, with coordinates $[50.0, -50.0, -100.0]$. The speed of sound $c = 345.0m/s$.

(unit: m)

		$\mathbf{M_1}$			$\mathbf{M_2}$			$\mathbf{M_3}$		
		x	y	z	x	y	z	x	y	z
Setup 1	Array 1	1.0	51.0	0.0	1.0	49.0	0.0	0.0	50.0	0.0
	Array 2	1.0	-49.0	0.0	1.0	-51.0	0.0	0.0	-50.0	0.0
	Pair 1	-49.0	1.0	0.0	-49.0	-1.0	0.0			
Setup 2	Pair 1	1.0	51.0	0.0	1.0	49.0	0.0			
	Pair 2	1.0	-49.0	0.0	1.0	-51.0	0.0			
	Pair 3	-49.0	1.0	0.0	-49.0	-1.0	0.0			
	Pair 4	51.0	1.0	0.0	-51.0	1.0	0.0			

Table A.1: This table shows the hydrophone position arrangement used in the simulation program. Setup1 consists of two hydrophone arrays and one hydrophone pair, and setup2 consists four hydrophone pairs.

Table A.2 shows the estimation results, The estimates from setup1 is very close to the real source position, with error percentage less than 0.01%. The estimates from setup2 is

less accurate, but still has the error percentage about 0.05%.

(unit: m)

	Estimates of Source Position		
	x	y	z
Setup 1	49.7055	-49.9717	-99.9852
Setup 2	46.6759	-49.8522	-101.5770

Table A.2: Simulation results in a perfect environment. The real source position is $[50.0, -50.0, -100.0]$

In order to investigate the performance of these two approaches, a random sampling error is added to the time delay estimation, and a random orientation error [1] is added to the estimate of the direction of arrival. The position of the sound source is measured 25 times with a random sampling error and orientation error added in a specific range, and the average estimate of the source position and standard deviation are calculated. Then the process is repeated at a new range of random errors. Figure A.3 shows the result.

[1] In the method using microphone arrays, the orientation error is obtained by rotating the direction vector calculated through Eq. (3.7) by an angle of random degree horizontally and then vertically; and in the method using hydrophone pair, the error is obtained by adding a random degree to the calculation of cosine and sine values in Eq. (A.3).

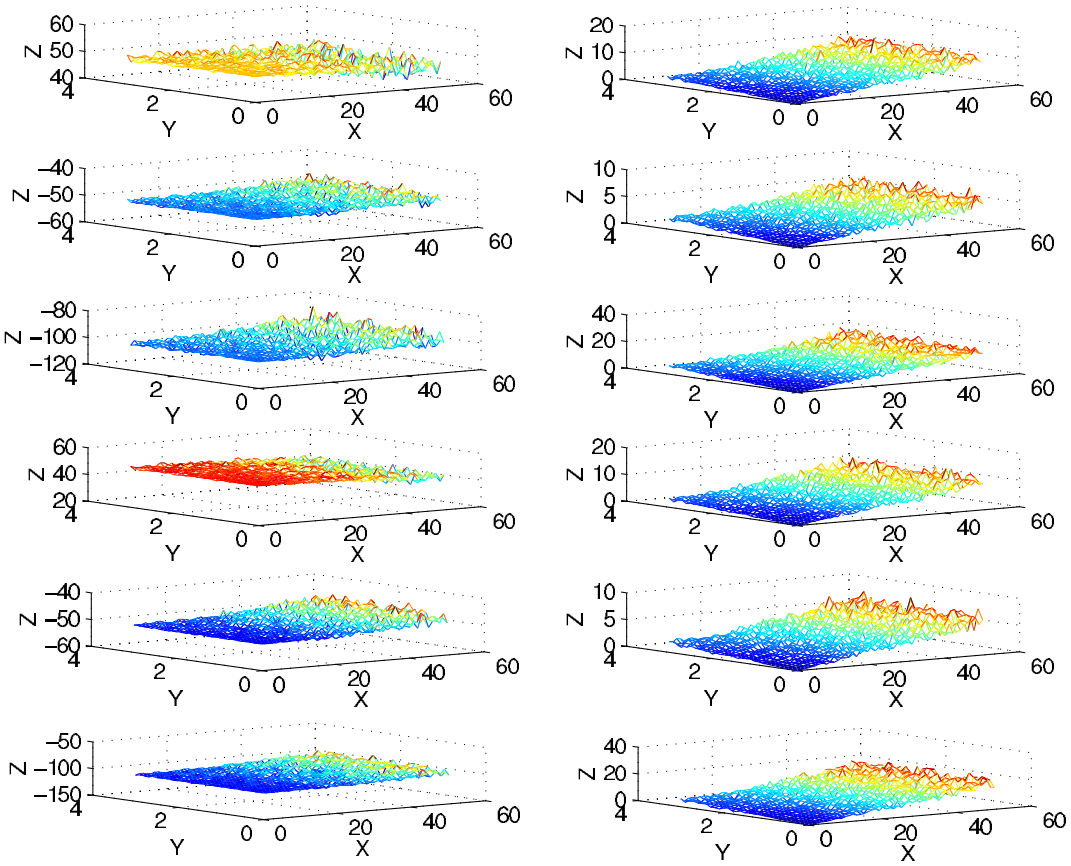

Figure A.3: Plots of the simulation results with orientation error and time estimation error added. The value along x dimension represents the range of random error (in samples) in time delay estimation. The value along y dimension represents the range of orientation error (in degrees). The top six plots are the results from Microphone Setup1, and the bottom six are the results from Setup2. The left are the Mean of the estimates along x, y and z dimensions. And the right are the STD of the estimates along x, y and z dimensions.

Appendix B

An Optimal Method for Position Measurement

The main tool that we use in our experiments to measure the absolute positions of sound source and microphone array positions is a measuring tape. We are interested in processing the measurements to extract positions as accurately as possible, given that different people may take different readings off the measuring tape, and even the same person may take different readings at different times. This appendix provides an optimal solution to reduce errors in the measurement of coordinates using a measuring tape. This includes two steps. The first is to measure the distances between pairs of points. The second step is to solve a minimization problem to find the coordinates of the points being measured. The problem is cast as a two-dimensional problem, by projecting all points to the ground plane using a plumbing bob.

First, we choose two points P_0 and P_1 to define the coordinate system, that is, P_0 as the origin, and P_1 as a point lying on the x-axis. Therefore, the coordinates of P_0 are $[0, 0]$, and the coordinates of P_1 are $[x_1, 0]$ where $y_1 = 0$. In order to reduce the estimation error for the remaining points, P_0 and P_1 must be far apart enough. Then the coordinate system can be determined by the right hand rule.

Consider another two points P_2 $[x_2, y_2]$ and P_3 $[x_3, y_3]$ the coordinates of which need to be determined. We measure the distances between each pair of points in the set $\{P_0, P_1, P_2, P_3\}$. Then the coordinates of these points can be obtained by solving the following minimization problem:

$$f = \min_{x_1,x_2,x_3,y_2,y_3} \sum_{k=0}^{n-1} \sum_{i,j=0;i\neq j}^{4} \left((x_i - x_j)^2 + (y_i - y_j)^2 - d_{ijk}^2\right) \tag{B.1}$$

where n is the number of the measurements, and d_{ijk} is the k-th measurement of the

distance between points P_i and P_j. This optimization problem can be solved numerically [12].

Appendix C

Raw Data from the Pool Experiment

In this appendix, we present the raw data from the pool experiment, where we combine multiple hydrophone array positions to simulate the availability of different numbers and positions of hydrophone array. At each array position, we collect the audio data, estimate the time delays between hydrophones, and then compute a direction line of the arrival of sound.

Table C.1 shows the measured raw delay in samples (including the outliers) at three hydrophones with respect to the remaining hydrophone in the array, where the array position is at $G1$, shown in Figure 5.17. The corresponding ideal delays at this position are $(-12, -6, 7)$, which are computed based on the knowledge of sound propagation model, and the measured positions of the hydrophones and the source.

The "robust" time delays (excluding the outliers) are shown in Table C.2. The method that we use for discarding outliers is described in Chapter 4.

The estimates of the sound direction vector corresponding to the delays in Table C.2 are shown in Table C.3, where the average of the estimates are:$[-0.03, 0.39, -0.92]$.

We calculate the average of the direction vectors at each hydrophone position and construct a direction line of the arrival of sound that starts at the center of the hydrophone array. Therefore, with a number of hydrophone array positions, we can construct multiple direction lines in 3D space.

Delay1	Delay2	Delay3	Delay1	Delay2	Delay3
-6	-10	4	-6	-11	3
-6	-10	4	-6	-10	4
-7	-10	3	-5	-11	3
-7	-11	3	-5	-13	4
-7	-10	3	-6	5	4
-6	-10	4	-5	-13	-13
-6	-10	4	-7	-11	4
-6	-10	4	-6	-11	4
-6	-9	4	-21	-23	3
-6	-11	3	-6	4	5
-6	-13	4	-7	-13	4
-6	-12	4	-7	-11	4
-7	-13	4	-7	-11	3
-5	-33	4	-7	-11	3
-7	-14	4	-7	-11	3
-7	-13	4	-6	-11	-13
-6	-12	4	-7	-11	2
-6	-11	4	-6	-11	2
-6	-11	3	-6	-11	2
-6	-11	3	-6	-11	2
-6	-11	4	-7	-11	2
-6	-13	3	-6	-11	2
-6	-12	3	-7	-12	2
-6	-12	4	-6	-12	2
-6	-12	4	-7	-11	2
-6	-13	4	-7	-11	3
-6	-11	5	-7	-12	2
-6	-11	5	-7	-11	3
-6	-12	4	-7	-11	2
-6	-12	4	-24	-12	1
-6	-13	4	-6	-12	1
-6	-12	4	-6	-12	0

Table C.1: The measured raw delays in integer units at three hydrophones with respect to the remaining hydrophone in the array of the pool experiment, which were estimated at position $G1$ shown in Figure 5.17.

Delay1	Delay2	Delay3	Delay1	Delay2	Delay3
-6	-10	4	-6	-10	4
-7	-10	3	-7	-11	3
-7	-10	3	-6	-10	4
-6	-10	4	-6	-10	4
-6	-9	4	-6	-11	3
-6	-12	4	-6	-12	4
-6	-11	4	-6	-11	3
-6	-11	3	-6	-11	4
-6	-12	3	-6	-12	4
-6	-12	4	-6	-11	5
-6	-11	5	-6	-12	4
-6	-12	4	-6	-12	4
-6	-11	3	-6	-10	4
-5	-11	3	-7	-11	4
-6	-11	4	-7	-11	4
-7	-11	3	-7	-11	3
-7	-11	3	-7	-11	2
-6	-11	2	-6	-11	2
-6	-11	2	-7	-11	2
-6	-11	2	-7	-12	2
-6	-12	2	-7	-11	2
-7	-11	3	-7	-12	2
-7	-11	3	-7	-11	2

Table C.2: The "robust" delays in integer units where the outliers were discarded from the delays in Table C.1.

x	y	z	x	y	z
0.0	0.38	-0.93	0.0	0.38	-0.93
0.0	0.38	-0.93	-0.02	0.39	-0.92
0.0	0.38	-0.93	0.0	0.38	-0.93
0.0	0.38	-0.93	0.0	0.38	-0.93
0.02	0.36	-0.93	-0.04	0.38	-0.93
-0.04	0.41	-0.91	-0.04	0.41	-0.91
-0.02	0.39	-0.92	-0.04	0.38	-0.93
-0.04	0.38	-0.93	-0.02	0.39	-0.92
-0.06	0.39	-0.92	-0.04	0.41	-0.91
-0.04	0.41	-0.91	0.0	0.41	-0.91
0.0	0.41	-0.91	-0.04	0.41	-0.91
-0.04	0.41	-0.91	-0.04	0.41	-0.91
-0.04	0.38	-0.93	0.0	0.38	-0.93
-0.06	0.36	-0.93	0.0	0.41	-0.91
-0.02	0.39	-0.92	0.0	0.41	-0.91
-0.02	0.39	-0.92	-0.02	0.39	-0.92
-0.02	0.39	-0.92	-0.04	0.38	-0.93
-0.06	0.36	-0.93	-0.06	0.36	-0.93
-0.06	0.36	-0.93	-0.04	0.38	-0.93
-0.06	0.36	-0.93	-0.06	0.39	-0.92
-0.08	0.38	-0.92	-0.04	0.38	-0.93
-0.02	0.39	-0.92	-0.06	0.39	-0.92
-0.02	0.39	-0.92	-0.04	0.38	-0.93

Table C.3: The estimates of the direction vectors calculated from the "robust" delays in Table C.2.

Appendix D

Experimental Results of Pool Experiment

In the pool experiment, we pair the direction lines to simulate pairs of hydrophone arrays, and then estimate the source location that has the minimum distance to the hydrophone arrays, as described in Sec. 3.3. Table D.1 and Table D.2 are the estimation results from the two kinds of sources, one generates sound by a person from the surface hitting a long metal pipe directly on the bottom of the pool, the other generates sound by a person who dives to the bottom hitting two pieces of metal pipe against each other. The actual source position is$[11.59, 7.91, -4.30]$, and the eight hydrophone array positions are represented as $G1 - G8$ which are shown in Figure 5.17.

We also combine multiple hydrophone array positions, with one direction line estimated at each position, thereby simulating different number and positions of hydrophone arrays. Table D.3 and Table D.4 shows the sound source position estimates based on three hydrophone positions.

(unit: m)

Hydrophone Pair	Source Position			Hydrophone Pair	Source Position		
	x	y	z		x	y	z
$G1, G2$	9.73	6.16	-0.08	$G3, G5$	12.46	11.38	-0.94
$G1, G3$	9.79	7.76	-0.19	$G3, G6$	12.09	10.16	-0.53
$G1, G4$	10.30	9.57	0.46	$G3, G7$	11.74	8.84	-0.21
$G1, G5$	11.32	8.98	-1.36	$G3, G8$	11.00	7.49	-0.49
$G1, G6$	11.38	7.77	-0.78	$G4, G5$	11.45	12.14	-0.11
$G1, G7$	11.50	6.41	0.03	$G4, G6$	11.38	11.06	0.34
$G1, G8$	11.13	5.07	-0.44	$G4, G7$	11.47	10.02	0.38
$G2, G3$	11.44	9.85	-0.40	$G4, G8$	11.50	8.97	0.72
$G2, G4$	11.43	11.19	0.03	$G5, G6$	12.60	10.96	-0.36
$G2, G5$	12.99	11.45	-0.73	$G5, G7$	12.60	9.65	-0.58
$G2, G6$	12.43	9.83	-0.43	$G5, G8$	12.36	8.32	-1.26
$G2, G7$	11.83	8.10	-0.11	$G6, G7$	12.56	8.60	-0.30
$G2, G8$	10.67	6.19	-0.49	$G6, G8$	12.44	7.33	-0.84
$G3, G4$	10.64	11.28	0.12	$G7, G8$	12.57	6.22	0.03

Table D.1: Estimates of the sound source position in the pool experiment based on pairs of hydrophone array positions, where the sound is generated by hitting two pieces of pipe against each other. The first and fifth column show the hydrophone array positions used for constructing pairs of sound direction lines. Column 2-4 and 6-8 are the corresponding estimates of the sound source position, in terms of their x, y and z coordinates.

(unit: m)

Hydrophone Pair	Source Position			Hydrophone Pair	Source Position		
	x	y	z		x	y	z
$G1, G2$	10.70	8.51	0.27	$G3, G5$	11.48	12.27	0.30
$G1, G3$	11.14	10.50	0.41	$G3, G6$	11.57	11.10	-0.15
$G1, G4$	11.54	12.81	0.67	$G3, G7$	11.61	9.74	-0.58
$G1, G5$	12.50	12.08	0.50	$G3, G8$	10.21	9.27	0.75
$G1, G6$	12.19	10.39	0.36	$G4, G5$	11.31	13.29	-0.10
$G1, G7$	11.87	8.50	0.25	$G4, G6$	11.13	12.26	-0.34
$G1, G8$	11.15	6.56	0.26	$G4, G7$	10.34	11.31	-0.33
$G2, G3$	11.06	10.41	-1.07	$G4, G8$	10.07	11.58	1.35
$G2, G4$	10.96	12.57	0.07	$G5, G6$	12.23	11.66	-0.02
$G2, G5$	12.15	12.10	0.12	$G5, G7$	11.06	10.61	0.20
$G2, G6$	11.93	10.71	-0.03	$G5, G8$	11.53	10.72	1.31
$G2, G7$	11.64	9.05	-0.11	$G6, G7$	11.71	9.79	0.08
$G2, G8$	10.41	7.70	0.58	$G6, G8$	11.82	9.42	1.00
$G3, G4$	10.17	12.76	0.53	$G7, G8$	12.21	7.74	0.53

Table D.2: Estimates of the sound source position in the pool experiment based on pairs of hydrophone array positions, where the sound is generated by striking a long pipe on the bottom. The first and fifth column represent the pairs of hydrophone array positions used in estimating pairs of sound direction lines. Column 2-4 and 6-8 are the corresponding estimates of the sound source position, in terms of their x, y and z coordinates.

(unit: m)

Hydrophone Pair	Source Position			Hydrophone Pair	Source Position		
	x	y	z		x	y	z
$G1, G2, G3$	9.87	7.70	-0.09	$G2, G4, G7$	11.60	9.89	0.27
$G1, G2, G4$	10.57	9.30	0.41	$G2, G4, G8$	11.41	9.03	0.51
$G1, G2, G5$	11.34	8.98	-0.85	$G2, G5, G6$	12.71	10.74	-0.62
$G1, G2, G6$	11.15	8.01	-0.43	$G2, G5, G7$	12.47	9.72	-0.53
$G1, G2, G7$	11.03	6.95	-0.03	$G2, G5, G8$	12.11	8.67	-0.77
$G1, G2, G8$	10.42	5.83	-0.43	$G2, G6, G7$	12.27	8.84	-0.28
$G1, G3, G4$	10.22	9.71	0.36	$G2, G6, G8$	11.93	7.82	-0.48
$G1, G3, G5$	11.19	9.46	-0.91	$G2, G7, G8$	11.78	6.89	-0.11
$G1, G3, G6$	11.05	8.62	-0.47	$G3, G4, G5$	11.49	11.67	-0.12
$G1, G3, G7$	10.99	7.71	-0.11	$G3, G4, G6$	11.30	10.88	0.24
$G1, G3, G8$	10.51	6.81	-0.50	$G3, G4, G7$	11.28	10.12	0.25
$G1, G4, G5$	11.00	10.46	0.05	$G3, G4, G8$	11.08	9.39	0.45
$G1, G4, G6$	10.98	9.64	0.46	$G3, G5, G6$	12.46	10.82	-0.78
$G1, G4, G7$	11.09	8.84	0.51	$G3, G5, G7$	12.29	9.96	-0.61
$G1, G4, G8$	11.08	8.04	0.73	$G3, G5, G8$	12.018	9.08	-0.89
$G1, G5, G6$	11.80	9.21	-1.19	$G3, G6, G7$	12.14	9.21	-0.34
$G1, G5, G7$	11.81	8.40	-0.61	$G3, G6, G8$	11.90	8.35	-0.57
$G1, G5, G8$	11.49	7.50	-1.26	$G3, G7, G8$	11.80	7.54	-0.19
$G1, G6, G7$	11.80	7.64	-0.28	$G4, G5, G6$	11.78	11.44	0.13
$G1, G6, G8$	11.59	6.75	-0.79	$G4, G5, G7$	11.82	10.70	0.08
$G1, G7, G8$	11.78	5.90	-0.03	$G4, G5, G8$	11.81	10.00	0.26
$G2, G3, G4$	11.26	10.89	0.04	$G4, G6, G7$	11.77	9.98	0.38
$G2, G3, G5$	12.66	11.12	-0.87	$G4, G6, G8$	11.80	9.27	0.63
$G2, G3, G6$	12.14	10.05	-0.52	$G4, G7, G8$	11.86	8.53	0.62
$G2, G3, G7$	11.61	8.89	-0.21	$G5, G6, G7$	12.62	9.69	-0.58
$G2, G3, G8$	10.74	7.65	-0.32	$G5, G6, G8$	12.53	8.79	-1.14
$G2, G4, G5$	11.96	11.65	-0.16	$G5, G7, G8$	12.53	8.08	-0.55
$G2, G4, G6$	11.73	10.76	0.20	$G6, G7, G8$	12.53	7.40	-0.27

Table D.3: Estimates of the sound source position in the pool experiment based three hydrophone array positions. The sound is generated by striking two short pieces of pipe against each other. The first and fifth column represent the hydrophone array positions used for constructing three sound direction lines. Column 2-4 and 6-8 are the corresponding estimates of the sound source position, in terms of their x, y and z coordinates.

(unit: m)

Hydrophone Pair	Source Position			Hydrophone Pair	Source Position		
	x	y	z		x	y	z
$G1, G2, G3$	10.97	9.91	-0.04	$G2, G4, G7$	10.63	10.94	-0.25
$G1, G2, G4$	11.44	12.08	0.50	$G2, G4, G8$	10.39	11.04	1.11
$G1, G2, G5$	12.18	11.67	0.49	$G2, G5, G6$	12.07	11.47	0.02
$G1, G2, G6$	11.86	10.36	0.30	$G2, G5, G7$	11.10	10.46	0.22
$G1, G2, G7$	11.49	8.83	0.16	$G2, G5, G8$	11.32	10.57	1.18
$G1, G2, G8$	10.66	7.62	0.44	$G2, G6, G7$	11.52	9.85	0.05
$G1, G3, G4$	11.03	12.33	0.74	$G2, G6, G8$	11.39	9.58	0.86
$G1, G3, G5$	11.81	11.88	0.56	$G2, G7, G8$	11.47	8.31	0.47
$G1, G3, G6$	11.68	10.72	0.18	$G3, G4, G5$	11.04	12.81	0.19
$G1, G3, G7$	11.46	9.41	-0.16	$G3, G4, G6$	10.96	12.03	-0.18
$G1, G3, G8$	10.82	8.93	0.48	$G3, G4, G7$	10.36	11.25	-0.29
$G1, G4, G5$	11.78	12.74	0.30	$G3, G4, G8$	9.90	11.51	1.33
$G1, G4, G6$	11.52	11.73	0.05	$G3, G5, G6$	11.75	11.66	0.02
$G1, G4, G7$	10.77	10.76	-0.03	$G3, G5, G7$	10.87	10.81	0.11
$G1, G4, G8$	10.95	11.24	1.31	$G3, G5, G8$	10.92	10.99	1.17
$G1, G5, G6$	12.25	11.32	0.26	$G3, G6, G7$	11.38	10.23	-0.13
$G1, G5, G7$	11.29	10.29	0.37	$G3, G6, G8$	11.15	10.05	0.70
$G1, G5, G8$	11.80	10.63	1.27	$G3, G7, G8$	11.38	8.90	0.15
$G1, G6, G7$	11.69	9.59	0.26	$G4, G5, G6$	11.51	12.38	-0.18
$G1, G6, G8$	11.80	9.44	0.97	$G4, G5, G7$	10.52	11.73	-0.09
$G1, G7, G8$	11.82	7.95	0.58	$G4, G5, G8$	10.94	11.91	0.85
$G2, G3, G4$	10.63	12.13	0.31	$G4, G6, G7$	10.86	11.15	-0.23
$G2, G3, G5$	11.51	11.83	0.31	$G4, G6, G8$	10.92	11.08	0.55
$G2, G3, G6$	11.51	10.90	-0.11	$G4, G7, G8$	10.41	10.28	0.39
$G2, G3, G7$	11.41	9.75	-0.51	$G5, G6, G7$	11.40	10.68	0.11
$G2, G3, G8$	10.33	9.08	0.36	$G5, G6, G8$	11.81	10.59	0.78
$G2, G4, G5$	11.52	12.73	0.03	$G5, G7, G8$	11.05	9.76	0.77
$G2, G4, G6$	11.32	11.86	-0.22	$G6, G7, G8$	11.61	9.09	0.61

Table D.4: Estimates of the sound source position in the pool experiment based on three hydrophone array positions. The sound is generated by striking a long pipe directly on the bottom of the pool. The first and fifth column represent the three hydrophone array positions used in constructing three sound direction lines. Column 2-4 and 6-8 are the corresponding estimates of the sound source position, in terms of their x, y and z coordinates.

Appendix E

A Sensitivity Analysis of Source Position Estimation

Our approach described in Chapter 3 uses far-field assumption to estimate a direction vector of the arrival of sound at a microphone array, and therefore determines the sound source position as the intersection of multiple direction lines pointing toward the source starting from the centers of multiple microphone arrays. In this appendix, we investigate the sensitivity of the error in the source position estimate with respect to the errors in the measurement of the sound direction vectors.

Consider two sound direction lines from two microphone arrays, with one starting at $\mathbf{P_1} = [p_{1x}, p_{1y}, p_{1z}$ and parallel to the direction vector $\mathbf{s_1} = [\theta 1, \phi 1]$, the other starting at $\mathbf{P_2} = [p_{2x}, p_{2y}, p_{2z}$ and parallel to the direction vector $\mathbf{s_2} = [\theta 2, \phi 2]$. Points $\mathbf{P_1}$ and $\mathbf{P_2}$ are the reference points (centers) of the two arrays. Vectors $\mathbf{s_1}$ and $\mathbf{s_2}$ are two unit vectors defined by spherical coordinates: azimuth and elevation. According to Section 3.3, we have a closed-form solution for the sound source position $\mathbf{SP} = [SP_x, SP_y, SP_z]$, which is computed by the following vector function:

$$\mathbf{SP} = f(\mathbf{P_1}, \mathbf{s_1}, \mathbf{P_2}, \mathbf{s_2}) \tag{E.1}$$

Eq. (E.1) can be viewed as being three scalar equations:

$$\begin{cases} SP_x = f_x(p_{1x}, p_{1y}, p_{1z}, \theta 1, \phi 1, p_{2x}, p_{2y}, p_{2z}, \theta 2, \phi 2) \\ SP_y = f_y(p_{1x}, p_{1y}, p_{1z}, \theta 1, \phi 1, p_{2x}, p_{2y}, p_{2z}, \theta 2, \phi 2) \\ SP_z = f_z(p_{1x}, p_{1y}, p_{1z}, \theta 1, \phi 1, p_{2x}, p_{2y}, p_{2z}, \theta 2, \phi 2) \end{cases} \tag{E.2}$$

where f_x, f_y, f_z are functions of ten variables. Small changes in the values of each of these ten variables results in a small change in the resulting coordinates of the source position. If

changes in the independent variables are sufficiently small, then the differential change in the resulting coordinate is approximated by a linear function of the changes in the variables. The proportionality constants are the the partial derivatives of SP_x, SP_y, SP_z with respect to each of the ten variables.

$$\begin{cases} \Delta SP_x = & c_{1x}\Delta p_{1x} + c_{2x}\Delta p_{1y} + c_{3x}\Delta p_{1z} + \\ & + c_{4x}\Delta\theta 1 + c_{5x}\Delta\phi 1 + c_{6x}\Delta p_{2x} + \\ & + c_{7x}\Delta p_{2y} + c_{8x}\Delta p_{2z} + c_{9x}\Delta\theta 2 + c_{10x}\Delta\phi 2 \\ \Delta SP_y = & c_{1y}\Delta p_{1x} + c_{2y}\Delta p_{1y} + c_{3y}\Delta p_{1z} + \\ & + c_{4y}\Delta\theta 1 + c_{5y}\Delta\phi 1 + c_{6y}\Delta p_{2x} + \\ & + c_{7y}\Delta p_{2y} + c_{8y}\Delta p_{2z} + c_{9y}\Delta\theta 2 + c_{10y}\Delta\phi 2 \\ \Delta SP_z = & c_{1z}\Delta p_{1x} + c_{2z}\Delta p_{1y} + c_{3z}\Delta p_{1z} + \\ & + c_{4z}\Delta\theta 1 + c_{5z}\Delta\phi 1 + c_{6z}\Delta p_{2x} + \\ & + c_{7z}\Delta p_{2y} + c_{8z}\Delta p_{2z} + c_{9z}\Delta\theta 2 + c_{10z}\Delta\phi 2 \end{cases} \tag{E.3}$$

where $c_{1x} \ldots c_{10x}$ are the partial derivatives of SP_x, $c_{1y} \ldots c_{10y}$ are the partial derivatives of SP_y, and $c_{1z} \ldots c_{10z}$ are the partial derivatives of SP_z, with respect to the ten variables p_{1x},p_{1y},p_{1z},$\theta 1$,$\phi 1$,p_{2x}, p_{2y},p_{2z},$\theta 2$,$\phi 2$.

Consider the setup in our pool experiment. We consider two hydrophone array positions, $G1$ and $G4$, as shown in Figure 5.17 and the estimated source position $[11.59, 7.91, -4.30]$. With these "ideal" positions, we have the corresponding "ideal" values for the variables p_{1x},p_{1y},p_{1z},$\theta 1$,$\phi 1$,p_{2x}, p_{2y},p_{2z},$\theta 2$,$\phi 2$ shown in Table E.1.

	Reference Point			Direction Vector	
	x	y	z	θ	ϕ
Array1	10.274	5.304	0.0	1.102	-0.975
Array2	10.274	12.254	0.0	-1.276	-0.759

Table E.1: This table shows the reference points of two hydrophone array positions, and the direction vectors pointing from these two arrays to the source position. They are measured from the pool experiment, and the angles are in radians.

The corresponding values of the partial derivatives of the functions in Eq. (E.2) with respect to these ten variables are shown in Table E.2.

Table E.2 reflects the impact of errors in the estimates of the sound direction vector on the estimates of the sound source positions.

Assume we have accurate measurement on the positions of reference points of hydrophone arrays. The error in the estimate of azimuth and elevation of the direction vectors can be computed by:

	∂SP_x	∂SP_y	∂SP_z		∂SP_x	∂SP_y	∂SP_z
$\partial\theta 1$	0.73	1.10	-1.53	$\partial\theta 2$	-0.97	0.96	2.47
$\partial\phi 1$	3.57	3.20	0.32	$\partial\phi 2$	5.30	-2.98	-3.02
$\partial p_1 x$	0.39	0.35	0.03	$\partial p_2 x$	0.61	-0.35	-0.04
$\partial p_1 y$	0.62	0.63	-0.19	$\partial p_2 y$	-0.62	0.37	0.19
$\partial p_1 z$	0.04	-0.10	0.49	$\partial p_2 z$	-0.04	0.11	0.51

Table E.2: This table shows the corresponding partial derivatives of the source position function with respect to the ten variables.

$$\Delta angle = angle_{ideal} - angle_{estimated} \tag{E.4}$$

Rewriting Eq. (E.3) becomes:

$$\begin{cases} |\Delta SP_x| &= |\frac{\partial SP_x}{\partial\theta 1}\Delta\theta 1 + \frac{\partial SP_x}{\partial\phi 1}\Delta\phi 1 + \frac{\partial SP_x}{\partial\theta 2}\Delta\theta 2 + \frac{\partial SP_x}{\partial\phi 2}\Delta\phi 2| \\ &\leq |\frac{\partial SP_x}{\partial\theta 1}\Delta\theta 1| + |\frac{\partial SP_x}{\partial\phi 1}\Delta\phi 1| + |\frac{\partial SP_x}{\partial\theta 2}\Delta\theta 2| + |\frac{\partial SP_x}{\partial\phi 2}\Delta\phi 2| \\ |\Delta SP_y| &= |\frac{\partial SP_y}{\partial\theta 1}\Delta\theta 1 + \frac{\partial SP_y}{\partial\phi 1}\Delta\phi 1 + \frac{\partial SP_y}{\partial\theta 2}\Delta\theta 2 + \frac{\partial SP_y}{\partial\phi 2}\Delta\phi 2| \\ &\leq |\frac{\partial SP_y}{\partial\theta 1}\Delta\theta 1| + |\frac{\partial SP_y}{\partial\phi 1}\Delta\phi 1| + |\frac{\partial SP_y}{\partial\theta 2}\Delta\theta 2| + |\frac{\partial SP_y}{\partial\phi 2}\Delta\phi 2| \\ |\Delta SP_z| &= |\frac{\partial SP_z}{\partial\theta 1}\Delta\theta 1 + \frac{\partial SP_z}{\partial\phi 1}\Delta\phi 1 + \frac{\partial SP_z}{\partial\theta 2}\Delta\theta 2 + \frac{\partial SP_z}{\partial\phi 2}\Delta\phi 2| \\ &\leq |\frac{\partial SP_z}{\partial\theta 1}\Delta\theta 1| + |\frac{\partial SP_z}{\partial\phi 1}\Delta\phi 1| + |\frac{\partial SP_z}{\partial\theta 2}\Delta\theta 2| + |\frac{\partial SP_z}{\partial\phi 2}\Delta\phi 2| \end{cases} \tag{E.5}$$

Table E.3 shows the estimates of the direction vectors from the two hydrophone array positions, whose "ideal" direction vectors are shown in Table E.1, and the errors between these estimates and the "ideal" directions.

$\theta 1$	-1.496	$\Delta\theta 1$	2.598
$\phi 1$	-0.999	$\Delta\phi 1$	0.024
$\theta 2$	1.475	$\Delta\theta 2$	-2.751
$\phi 2$	-1.075	$\Delta\phi 2$	0.316

Table E.3: This table shows the estimates of the direction vectors from two hydrophone array positions in the pool experiment and the errors between these estimates and the "ideal" sound directions. The angles are in radians.

Therefore, taking the errors in Table E.3, we can theoretically compute the upper bound of the errors in the sound position estimates through Eq. (E.5). Table E.4 shows the results.

(unit: m)

$max\Delta SP_x$	$max\Delta SP_y$	$max\Delta SP_z$
6.326	6.520	11.732

Table E.4: This table shows the theoretical upper-bound of the errors in the sound position estimates of the pool experiment.

Appendix F

Technical Specifications of Equipment

1. **Layla24 Digital Audio Recording**
   Manufacturer: Echo Technology (`http://www.echoaudio.com`)
   Model: 24
   Weight: $\approx 7lbs$
   Dimension: $30.0 \times 15.0 \times 4.0cm$ (length/width/height)
   Cost: $\approx US\$870$

2. **Omnidirectional Tie Clip Microphone**
   Manufacturer: Genexxa
   Model: 33-3003
   Weight: $\approx 20g$
   Dimensions: $17.6 \times 8.0mm$(length/diameter)
   Cost: $\approx US\$30$

3. **DolphinEar/Pro Omnidirectional Hydrophone**
   Manufacturer: Doyle-Renney Research (`http://www.dolphinear.com`)
   Model: Professional
   Weight: $\approx 1kg$ (including $20m$ cable)
   Dimensions: $60.0mm \times 8.0mm \times 20.0m$ (diameter/thick/cable length)
   Cost: $\approx US\$360$