Eye Array Sound Source Localization

by

Hedayat Alghassi

B.Sc., Sharif University of Technology, 1989

M.Sc., Sharif University of Technology, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Electrical and Computer Engineering)

UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

April 2008

© Hedayat Alghassi, 2008.

Abstract

Sound source localization with microphone arrays has received considerable attention as a means for the automated tracking of individuals in an enclosed space and as a necessary component of any general-purpose speech capture and automated camera-pointing system. A novel computationally efficient method compared to traditional source localization techniques is proposed and is both theoretically and experimentally investigated in this research.

This thesis first reviews the previous work in this area. The evolution of a new localization algorithm accompanied by an array structure for audio signal localization in three dimensional space is then presented. This method, which has similarities to the structure of the eye, consists of a novel hemispherical microphone array with microphones on the shell and one microphone in the center of the sphere. The hemispherical array provides such benefits as 3D coverage, simple signal processing and low computational complexity. The signal processing scheme utilizes parallel computation of a special and novel closeness function for each microphone direction on the shell. The closeness functions have output values that are linearly proportional to the spatial angular difference between the sound source direction and each of the shell microphone directions. Finally by choosing directions corresponding to the highest closeness function values and implementing linear weighted spatial averaging in those directions we estimate the sound source direction. The experimental tests validate the method with less than 3.1° of error in a small office room.

Contrary to traditional algorithmic sound source localization techniques, the proposed method is based on parallel mathematical calculations in the time domain. Consequently, it can be easily implemented on a custom designed integrated circuit.

Table of Contents

ABSTRACT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
LIST OF ABBREVIATIONS	xii
ACKNOWLEDGMENT	xiv
DEDICATION.....	xv
1. INTRODUCTION	1
1.1 MOTIVATION	2
1.2 TYPICAL APPLICATION: CAMERA POINTING	3
1.3 GOAL AND CONTRIBUTIONS OF THE THESIS	4
1.3.1 Goal	4
1.3.2 Contributions of the Thesis	4
1.4 ORGANIZATION OF THE THESIS.....	5
2. BACKGROUND: EXISTING METHODS	7
2.1 SOURCE LOCALIZATION STRATEGIES.....	7
2.1.1 Steered Beamformer Based Approaches.....	8
2.1.2 High Resolution Spectral-Estimation Based Locators.....	10
2.1.3 Time Delay Of Arrival (TDOA) Based Locators.....	11
2.2 TDOA BASED SOURCE LOCATORS.....	13
2.2.1 Time Delay Estimation.....	13
2.2.2 Cross-Power Spectrum Phase (CPSP).....	16

2.2.3 Modified CPSP	17
2.2.4 Source Location Finding Algorithm.....	18
2.3 ADVANCED DEVELOPMENTS IN SOUND SOURCE	
LOCALIZATION.....	22
2.3.1 General.....	22
2.3.2 Distributed Arrays.....	22
2.3.3 Multi Rate, Spectral, Non-free Field.....	23
2.3.4 Orientation Assisted.....	25
2.3.5 Optimization Based.....	26
2.3.6 TDOA Enhancement.....	27
2.3.7 Pre-filtering and Clustering.....	28
2.4 ORTHOGONAL MICROPHONE WORK.....	29
2.5 SUMMARY	30
3. EYE ARRAY IN SOUND SOURCE LOCALIZATION.....	32
3.1 A NEW LOCALIZATION METHOD	32
3.2 ASSUMPTIONS	33
3.3 OVERVIEW	34
3.3.1 Hemisphere versus Sphere.....	34
3.3.2 Method Synopsis	35
3.4 DEFINITIONS	37
3.5 TWO-MICROPHONE CELL	38
3.6 SPHERICAL TRAVEL TIME DELAY	42
3.6.1 Derivative Sampling.....	43
3.6.2 Derivation Time.....	45
3.6.3 Processing Time Frame and Noise.....	45
3.7 THREE-MICROPHONE CELL	47
3.8 ARRAY TOPOLOGY	49
3.9 ALGORITHM.....	50
3.9.1 Formulation.....	51
4. ALTERNATIVE CLOSENESS FUNCTIONS	53
4.1 DIFFERENCE CLOSENESS FUNCTION (DCF).....	53
4.1.1 Difference Two-Microphone (Pinhole) CF.....	53
4.1.2 Difference Three-Microphone (Lens) CF	57
4.1.2.1 Pseudo Transfer Function	59
4.2 CORRELATIVE CLOSENESS FUNCTION (CCF)	64

4.2.1 Pinhole and Lens Correlative CF	64
4.3 JOINED CLOSENESS FUNCTIONS	67
4.4 COROLLARIES	68
5. EXPERIMENTAL RESULTS.....	70
5.1 TEST PROTOCOLS	70
5.1.1 Environment	71
5.1.2 Sound Source, Acquisition	72
5.1.3 Visualization	74
5.1.4 Error Calculation	75
5.2 RESULTS UTILIZING MCF	76
5.3 RESULTS UTILIZING DCF	85
5.4 RESULTS UTILIZING CCF	90
5.5 EFFECT OF THE SIGNAL-TO-NOISE RATIO.....	95
5.6 ARRAY SIZE, SYSTEM NOISE, OUTLIERS, NUMBER OF CELLS	97
5.7 REMARKS.....	99
6. REFINEMENT OF THE APPROACH: EYE ARRAY	
PLACEMENT.....	100
6.1 WHERE TO PLACE THE ARRAY?.....	100
6.2 LENS CELL SENSITIVITY TO REVERBERATION	102
6.3 TRIHEDRAL CORNER RETROREFLECTION PROPERTY	104
6.4 UPPER TRIHEDRAL CORNER PLACEMENT	106
6.5 EVALUATION	107
6.6 MULTIPLE ARRAY LOCALIZATION AND PLACEMENT	109
7. IMPLEMENTATION ISSUES	111
7.1 PRELIMINARY TEST BENCH.....	111
7.2 MECHANICAL DESIGN AND CONSTRUCTION	113
7.2.1 Topology Selection.....	114
7.2.2 Array Design	117
7.2.3 Mechanical Structure.....	118
7.3 ELECTRICAL IMPLEMENTATION	123

7.3.1 Microphones	123
7.3.2 Data Acquisition Board	126
7.3.3 Preamplifiers.....	127
7.3.4 Computer.....	128
7.3.5 Software	130
7.3.6 Computational Complexity	132
7.4 METHOD OF CHOICE: INTEGRATED CIRCUIT	137
8. CONCLUSIONS	140
8.1 SUMMARY OF THESIS CONTRIBUTIONS	140
8.2 DISADVANTAGES	142
8.3 FUTURE DIRECTIONS AND ENHANCEMENTS.....	143
8.3.1 Perspective.....	143
8.3.2 Approach	144
REFERENCES	147
APPENDIX A.....	161
DIGITAL FORMULATION OF THE MCF ALGORITHM.....	161
MCF Algorithm Flowchart.....	164
APPENDIX B.....	166
FAR FIELD ASSUMPTION	166
APPENDIX C.....	168
EYE ANALOGY	168

List of Tables

TABLE 5-1 ERROR AND COVERAGE VERSUS CLOSENESS FUNCTION	99
TABLE 7-1: EYE ARRAY SHELL MICROPHONE DIRECTIONS WITH THEIR ORTHOGONAL PAIRS.....	115

List of Figures

FIGURE 1-1: SOUND SOURCE LOCALIZATION PROBLEM IN AN ENCLOSED AREA.....	2
FIGURE 2-1: DELAY-AND-SUM BEAMFORMER.....	8
FIGURE 2-2 TIME DELAY ESTIMATION OF A MICROPHONE PAIR.....	12
FIGURE 2-3 TWO-STEP TDOA BASED LOCALIZATION.....	13
FIGURE 2-4 SOURCE-SENSOR SETTING IN TDE	14
FIGURE 2-5 SOURCE LOCUS IN 2D [19]	19
FIGURE 2-6 SOURCE LOCUS IN 3D, HYPERBOLOID HYPER PLANE [4]	20
FIGURE 2-7 SOURCE LOCUS IN 3D CONE APPROXIMATION OF HYPERBOLOID [4]	21
FIGURE 3-1 A SPHERICAL TRIANGLE COVERAGE BY THREE CLOSENESS FUNCTIONS	36
FIGURE 3-2. REPRESENTATIONS IN SPHERICAL COORDINATES.....	39
FIGURE 3-3 VISUAL REPRESENTATION OF TWO-MICROPHONE COSINE SHAPED CLOSENESS FUNCTIONS	42
FIGURE 3-4 CONTINUOUS (LEFT) AND QUANTIZED (RIGHT) TRAVEL TIME FROM THE REFERENCE MICROPHONE TO A SPHERICAL SHELL	44
FIGURE 3-5 TWO-FREQUENCY ICOSAHERAL GEODESIC HEMISPHERE MICROPHONE ARRAYS.....	50
FIGURE 4-1 TWO-MICROPHONE DIFFERENCE CONFIGURATION	54
FIGURE 4-2 FREQUENCY VERSUS ANGLE RESPONSE OF DIFFERENCE ELEMENT.....	56
FIGURE 4-3 ANGULAR RESPONSE OF INTEGRATED DIFFERENCE ELEMENT	57
FIGURE 4-4 LENS MICROPHONE CELL	58
FIGURE 4-5 PSEUDO FREQUENCY VERSUS ANGLE RESPONSE OF DIFFERENCE LENS CF ..	61

FIGURE 4-6 ANGULAR RESPONSE OF INTEGRATED LENS CELL.....	62
FIGURE 5-1. THE EYE MICROPHONE ARRAY UNDER TEST.....	72
FIGURE 5-2 CLOSENESS FUNCTION VISUALIZATION WITH MICROPHONE NUMBERING ...	75
FIGURE 5-3 CLOSENESS FUNCTION SNAPSHOT (X IS THE SOUND SOURCE DIRECTION)...	78
FIGURE 5-4 FOUR DIFFERENT SNAPSHOT OF MCF WITH SOURCE DIRECTIONS	79
FIGURE 5-5 THE MCF ERROR VERSUS AZIMUTH AND ELEVATION.....	80
FIGURE 5-6 QUIVER PLOT OF MCF GRADIENT ERROR VERSUS AZIMUTH AND ELEVATION	81
FIGURE 5-7 QUIVER PLOT OF CENTRAL AREA OF THE MCF GRADIENT.....	83
FIGURE 5-8 MCF VERTICAL AND HORIZONTAL ERROR PLOT.....	85
FIGURE 5-9 THE DCF ERROR VERSUS AZIMUTH AND ELEVATION.....	86
FIGURE 5-10 QUIVER PLOT OF DCF GRADIENT ERROR VERSUS AZIMUTH AND ELEVATION.....	87
FIGURE 5-11 A SAMPLE SNAPSHOT OF DCF VALUES WITH VISIBLE OUTLIERS.....	88
FIGURE 5-12 QUIVER PLOT OF CENTRAL AREA OF DCF GRADIENT	88
FIGURE 5-13 DCF VERTICAL AND HORIZONTAL ERROR PLOT.....	89
FIGURE 5-14 THE CCF ERROR VERSUS AZIMUTH AND ELEVATION	91
FIGURE 5-15 QUIVER PLOT OF CCF GRADIENT ERROR VERSUS AZIMUTH AND ELEVATION.....	92
FIGURE 5-16 QUIVER PLOT OF CENTRAL AREA OF CCF GRADIENT.....	93
FIGURE 5-17 CCF VERTICAL AND HORIZONTAL ERROR PLOT	94
FIGURE 5-18 AVERAGE ERROR PERFORMANCE VERSUS SNR.....	96

FIGURE 6-1 KIRCHER’S ACOUSTICAL PERCEPTION; THE EMERGENCE OF REFLECTION AND ECHOES [40].....	101
FIGURE 6-2 LENS CELL WITH FOREMOST REFLECTION	102
FIGURE 6-3 ORTHOGONAL TRIHEDRAL CORNER	105
FIGURE 6-4 EYE ARRAY PLACEMENT IN UPPER TRIHEDRAL CORNER.....	107
FIGURE 6-5 RMS ERROR VERSUS SOURCE BEARING ANGLE FOR DIHEDRAL CORNER AND SINGLE WALL PLACEMENTS.....	108
FIGURE 7-1: PRELIMINARY TEST BENCH	112
FIGURE 7-2 THE PINHOLE (TOP) AND LENS (BOTTOM) NORMALIZED MEASURED DCF VERSUS ANGLE	113
FIGURE 7-3 WIREFRAME FRONTAL VIEW OF ARRAY	114
FIGURE 7-4 SOLID REPRESENTATION OF THE TOPOLOGY []	116
FIGURE 7-5 EYE ARRAY DESIGN VIEW	117
FIGURE 7-6 THE CONSTRUCTED ARRAY STRUCTURE	118
FIGURE 7-7 MICROPHONE JOINT HOLDERS	119
FIGURE 7-8 ARRAY GONIOMETERS.....	120
FIGURE 7-9 FIXED COORDINATE AT CORNER VERSUS MOVABLE COORDINATE AT CENTER OF HEMISPHERE.....	121
FIGURE 7-10 MICROPHONE SHAPE, DIRECTIVITY PATTERN, CONNECTION AND FREQUENCY RESPONSE.....	124
FIGURE 7-11 MEASUREMENT CIRCUIT FOR THE BSE MICROPHONE	125
FIGURE 7-12 FREQUENCY RESPONSE OF BSE MICROPHONE	125
FIGURE 7-13 NATIONAL INSTRUMENT PCI-6071E DATA ACQUISITION BOARD	126

FIGURE 7-14 CABLE AND CONNECTOR BLOCK	127
FIGURE 7-15 PREAMPLIFIER/FILTER DAUGHTER BOARDS	128
FIGURE 7-16 THE LOUDSPEAKER ON STAND WITH THE ATTACHED LASER POINTER	129
FIGURE 7-17 BRUEL & KJAER INTEGRATING SOUND LEVEL METER.....	130
FIGURE 7-18 THE MCF EXPERIMENT PLATFORM IN LABVIEW	131
FIGURE 7-19 THE MCF ALGORITHM.....	132
FIGURE 7-20 COMPUTATIONAL PERFORMANCE COMPARISON BENCHMARKS.....	135
FIGURE 8-1 THEORETICAL (LEFT) AND EXPERIMENTAL (RIGHT) COVERAGE.....	140
FIGURE A-1 FLOWCHART OF THE MCF BASED EYE ARRAY PROCESSING	165
FIGURE B-1 SPHERICAL RADIATION OF SPHERICAL SOUND SOURCE.....	166
FIGURE C-1 EYE MECHANISM [119].....	169

List of Abbreviations

TDOA: Time Delay Of Arrival

TDE: Time Delay Estimation

DOA: Direction Of Arrival

GCC: Generalized Cross Correlation

SRP: Steered Response Power

PHAT: Phase Transform

CPSP: Cross Power Spectrum Phase

DFT: Discrete Fourier Transform

FFT: Fast Fourier Transform

SL: Source Localization

SSL: Sound Source Localization

CF: Closeness Function

MCF: Multiplicative Closeness Function

CCF: Correlative Closeness Function

DCF: Difference Closeness Function

SNR: Signal-to-Noise Ratio

ISLM: Integrating Sound Level Meter

3D: Three Dimensional

RT: Reverberation Time

LMS: Least Mean Square

ML: Maximum-Likelihood

MV: Minimum Variance

MA: Moving Average

BMA: Block Moving Average

MUSIC: Multiple Signal Classification

ESPRIT: Estimation of Signal Parameters via Rotational Invariance Techniques

RMS: Root Mean Square

DAQ: Data Acquisition

IC: Integrated Circuit

A/D: Analog to Digital converter

EMI: Electro-Magnetic Interference

RFI: Radio Frequency Interference

BSE®: Best Sound Electronics

NI®: National Instruments

Acknowledgment

I wish to express my sincere gratitude to my supervisor Professor Peter Lawrence for his great insight, mentoring and feedback during the course of this project. Also thanks to the research cosprovider Dr. Shahram Tafazoli for his friendship and great feedback on this thesis. Thanks also for the constructive feedback and corrections from the readers and university examiners Professor Murray Hodgson, Dr. Shahriar Mirabbasi and Dr. Sidney Fels, also the external examiner Professor Les Atlas.

I also thank Mr. Simon Bachman for his help on mechanical construction of the test array structure. Thanks also to my brother Mr. Hamid Alghassi for his help on drawing the three dimensional pictures on this thesis.

I would also thank my mother, all of my brothers and sisters for their endless support. My fiancée's understanding and support over the years have been the reason I have been able to complete this goal. I appreciate her more than she will ever know.

Dedication

Dedicated to the memory of my dearest father.

1. Introduction

The localization of sources of emitting signals has been the focus of attention for more than a century. The two dominant areas of application for such systems and algorithms are traditionally radar and underwater acoustics. Localization of sound sources in the medium of air via microphones is fairly new compared to those applications. Arrays of microphones have numerous advantages compared to single microphone systems. Localization and aiming in addition to noise and interference rejection allow microphone arrays to outperform single microphone systems.

Arrays of microphones have a variety of applications in speech data acquisition systems. Applications include teleconferencing, speech recognition, speaker identification, sound capture in adverse environments, biomedical devices for hearing impaired, audio surveillance, gunshot detection and camera pointing systems.

The fundamental requirement for sensor array systems is the ability to locate and track a signal source. In addition to high accuracy, the location estimator must be capable of a high update rate at reasonable computational load in order to be useful for real time tracking and beamforming applications. Source location data may also be used for purposes other than beamforming, such as aiming a camera in a video conferencing system (Figure 1-1).

This thesis uses the term “localization” to mean either a) direction (i.e. bearing) and range or b) direction only. This practice is also common in the related literature [1, 2].

The technique proposed in this thesis addresses only the estimation of source bearing.

This research addresses the specific applications of source localization methods for estimating the direction of a sound source in an enclosed environment to enable three dimensional (3D) applications, given limited computational resources.

Figure 1-1: Sound source localization problem in an enclosed area

1.1 Motivation

Most of the present sound source localization systems assume that the sound sources are distributed in a horizontal plane. This assumption simplifies the problem of sound source localization in almost all previous methods. In teleconference applications they assume all talkers speak at the same height which is somewhat true, but the talker or other attendees can act as sound blockades between the main talker and the array, which is typically a linear wall-mounted microphone array.

In most dominant sound source localization methods, the computational cost for two dimensional cases is high so that the real time implementation needs a computer with

high processing power. Some of these sound source localization methods have been modified to cover a three dimensional space at a very high computational cost.

There is thus a need for a sound source localization technique in 3D space that can be implemented in real time without requiring high computational power.

1.2 Typical Application: Camera Pointing

In most of the current video conferencing systems, speakers are often constrained to a small range in front of the camera, since the use of a wide-angle camera reduces image resolution. In some current video-conferencing systems, a collection of human-controlled cameras are set up in various locations to provide video of the active speaker as he or she contributes to the conference. Usually, this tiresome task needs full participation of the professional camera operators. It would certainly be better if the camera automatically frames any speaker, and permits the speaker to walk freely around the room. In order to do so, one would have to detect the location of the speaker and control the direction of the camera in real time. Acoustic and image object tracking techniques can be used to locate and track an active speaker automatically in 3D space, estimate his or her azimuth, elevation and range. Afterward, the need for human camera operators can be eliminated and the usage of the system will automatically increase.

There are several methods for tracking an active talker. These methods can be generally classified as either visual tracking or acoustic tracking, depending on the availability and usage of the particular cue (visual or acoustic).

Even though visual tracking techniques have been investigated for several decades and have had good success along with very high computation time, acoustic source localization systems have some advantages that are not present in vision-based tracking

systems. For example, they are computationally more efficient than visual systems. Furthermore, they receive acoustic signals omni-directionally and can act in the darkness. Therefore, they are able to detect and locate sound sources in the rear or sources that are hiding or occluded.

Humans, like most vertebrates, have two ears, which form a two microphone array, mounted on a mobile base that is the head. By continuously receiving and processing the propagating acoustic signals with such a binaural auditory system, humans can precisely and instantly gather information about the environment, particularly about the spatial positions and trajectories of sound sources and about their state of activity. However, the extraordinary performance features achievable by the binaural auditory system are a big technical challenge for engineers to reproduce, mainly because of the room reverberation, background noise and computational complexity.

1.3 Goal and Contributions of the Thesis

1.3.1 Goal

The goal of this thesis is to develop a computationally efficient and accurate means to estimate the 3D direction of a single dominant sound source located in an enclosed three dimensional space.

1.3.2 Contributions of the Thesis

To date there has been no work on sound source localization, to the best of our knowledge, that employs a 3-microphone orthogonal cell element, or employs a hemispherical array of microphones with a microphone in the center, or uses a fully time

domain computation to estimate a wideband sound source direction as proposed in this thesis.

By examining the previous literature (Chapter 2), we also demonstrate that previous approaches require greater computational load than the method in this thesis.

Briefly, this thesis makes the following contributions:

- A new theoretical approach to determine the direction of a sound source by using a special array structure to subdivide a 3D search area into complementary and parallel sections of equal problems, as described later in this thesis.
- A novel microphone array structure
- Theoretical analysis of a variety of closeness functions devised for the proposed structure
- Experimental verification justifying the approach and quantifying the error, coverage, noise analysis and computational complexity

In terms of accuracy, since there is no common experimental test-bed employed by all researchers it is difficult to definitively compare any absolute accuracy claims with our system accuracy.

1.4 Organization of the Thesis

We initially focus on the background by reviewing some of the dominant sound source localization methods available to date in Chapter 2. Next, in chapter 3, we discuss the theoretical framework of our new sound localization technique followed by some of the results. The concept of an eye array and closeness function is described in this chapter. In Chapter 4 we present two alternative forms for closeness functions followed by the

results of our system with those closeness functions. Chapter 5 deals with a novel placement strategy for our eye array which reduces the effect of reverberation in enclosed areas. Chapter 6 deals with all of implementation related aspects of our work including mechanical and electrical aspects and software details. Chapter 7 concludes this thesis with achievements and drawbacks of our work and later provides directions for future related research.

2. Background: Existing Methods

In this chapter, we first briefly introduce three different categories of sound source localization strategies. Later, we elaborate on the time delay of arrival based strategy and discuss it in depth. Finally, we discuss the recent advances in this research area followed by some remarks.

2.1 Source Localization Strategies

Existing source localization methods can be loosely divided into three categories: those based upon maximizing the steered response power of a beamformer, high-resolution spectral estimation-based, and time delay estimation based locators [1, 2, 3, 4].

- The first category corresponds to those methods in which the location estimate is derived directly from a filtered, weighted and summed version of the signal data received at the sensors.
- The second category represents any localization scheme relying upon an application of the correlation matrix of the signal.
- The third category includes procedures, which calculate source locations from a set of delay estimates measured across various combinations of the microphones.

With continued investigation over the last decade, the time delay estimation based locator has become the technique of choice, especially in recent digital systems due to its higher accuracy and comparably lower computational complexity [1, 2, 3, 4]. Next, each of these methods is discussed in detail.

2.1.1 Steered Beamformer Based Approaches

A steered beamformer steers a microphone array to various locations and searches for a peak in the output power, named focalization. The simplest type of steered response is obtained using the output of a delay-and-sum beamformer. This method is often referred to as a conventional beamformer. Delay and sum beamformers apply time shifts to the array signals to compensate for the propagation delays in the arrival of the source signal at each microphone (Figure 2-1). These signals are time aligned and summed together to form a single output signal.

Figure 2-1: Delay-and-sum beamformer

More complicated beamformers apply filters to the array signals as well as the time alignment. Derivation of filters in these filter-and-sum beamformers distinguishes one variety from another.

Beamforming has been extensively used in speech array applications for voice capture. However, due to the efficiency and satisfactory performance of the other methods, it has not often been applied to the source localization problem.

The optimal Maximum Likelihood (ML) estimator is used to focus a beamformer. The physical realization of ML estimator requires the solution of a non-linear optimization problem. The use of standard iterative optimization methods, such as the steepest descent and the Newton-Raphson method, for these processes was addressed in [5]. The main drawback of these approaches is that the cost function, which has to be maximized, does not have a strong global peak and frequently contains several local maxima. As a result, this kind of efficient search method is often inaccurate and extremely sensitive to the initial search location.

Overall, the computational requirements of the focalization based ML estimator, comprised of a complex objective function as well as the relative inefficiency of an appropriate optimization procedure, prohibit its use in the majority of practical, real time source direction estimators. Furthermore, the steered response of a conventional beamformer is highly dependent on the spectral content of the source signal. Many optimal derivations are based on *a priori* knowledge of the spectral content of the background noise as well as of the source signal [6, 7]. In the presence of significant reverberation, the noise and source signals are highly correlated which makes accurate estimation of the noise infeasible. Furthermore, in virtually all array applications, little or nothing is known about the source signal. Hence, such optimal estimators are not very practical in realistic speech array environments [1].

2.1.2 High Resolution Spectral-Estimation Based Locators

This second category of location estimation techniques includes the modern beamforming methods adapted from the field of high-resolution spectral analysis, minimum variance (MV) spectral estimation and the variety of eigen-analysis-based techniques (e.g. MUSIC and ESPRIT) [8]. Detailed description of these approaches can be found in [9]. While these techniques have successfully found their way into a variety of array processing applications, they all possess certain restrictions that have been found to limit their effectiveness in the sound source localization problem.

Each of these high-resolution processes is based upon the spatio-spectral correlation matrix derived from the signals received at the sensors. When exact knowledge of this matrix is unknown (which is almost always the case), it must be estimated from the observed data. This is done via ensemble averaging of the signals over an interval in which the sources and noise are assumed statistically stationary and their estimation parameters (location) are assumed to be fixed. For speech signals, fulfilling these conditions while allowing sufficient averaging can be very problematic in practice. With regard to the localization problem at hand, these methods were developed in the context of far field plane waves projecting onto a linear array. While the MV, MUSIC and ESPRIT algorithms have been shown to be extendible to the case of general array geometries and near-field sources [10], certain eigen-analysis approaches are limited to the far-field, uniform linear array situation.

With regard to the issue of computational load, a search of the location space is required in each of these scenarios. While the computational complexity at each iteration is not as demanding as in the case of steered beamformer, it is often too high to be simply realized in real time. Moreover, the objective space typically consists of sharp peaks. This

property prevents the use of iteratively efficient optimization methods. It should be noted that these high-resolution methods are all designed for narrowband signals. They can be extended to wideband signals, including speech, through simple serial application of narrowband methods or more sophisticated generalizations of these approaches [11.. Either of these methods extends the computational requirements considerably.

These algorithms tend to be significantly less robust to source and sensor-modeling errors than conventional beam forming methods [12]. The incorporated models typically assume ideal source radiators, uniform sensor channel characteristics, and exact knowledge of sensor positions. Such conditions are impossible to obtain in real environments. While the sensitivity of these high-resolution methods to the modeling assumptions may be reduced, it is at the cost of performance. Additionally, signal coherence, such as that created by the reverberation, is detrimental to algorithmic methods particularly that of eigenanalysis approaches. For these reasons, sound source localization methods based on high-resolution strategies have not been considered recently, except for some multi source situations [1, 2].

2.1.3 Time Delay Of Arrival (TDOA) Based Locators

Time delay estimation is concerned with the computation of the relative time delay of arrival between different microphone sensors. It is a fundamental technique in microphone array signal processing and the first step in passive TDOA based acoustic source localization systems. With this kind of localization, a two-step strategy is adopted (Figure 2-3). Time delay estimation (TDE) of the speech signals relative to pairs of microphones is performed first (Figure 2-2).

Figure 2-2 Time delay estimation of a microphone pair

This data along with knowledge of the microphone positions are then used to generate hyperbolic curves, which are then intersected in some optimal sense to arrive at a source location estimate. Several variations of this principle have been developed [13]. They differ considerably in the method of derivation, the extent of their applicability (2D versus 3D, near field source versus far field source), and their means of solution. Primarily because of their computational practicality and reasonable performance under good conditions, the bulk of passive speech localization systems in use today are TDOA based. Accurate and precise TDE is the key to the effectiveness of localizers within this group. The two major sources of signal degradation, which complicate this estimation problem, are background noise and channel multipath due to room reverberations.

Figure 2-3 Two-step TDOA based localization

2.2 TDOA Based Source Locators

Here we will describe the time delay of arrival estimation based methods with more emphasis on cross power spectrum phase time delay estimation, as the main method for most of the current sound source localization research. Extensive description of this method compared to the others is due to the above mentioned fact.

2.2.1 Time Delay Estimation

This method was first introduced by Omologo and Savizer [14, 15, 16], based on phase based time delay estimation of sensor pairs [17] and generalized cross correlation[18] and later modified by other researchers [1, 2, 4].

Given a single source of sound that produces a time varying signal $x(t)$, each microphone in the array will receive the signal:

$$m_i(t) = \alpha_i x(t - \Delta_i) + n_i(t); \quad \Delta_i = \frac{D_i}{V_{\text{sound}}} \quad (2.1)$$

Figure 2-4 Source-sensor setting in TDE

Where i is the microphone number, Δ_i is the time it takes for the sound to propagate from the source S to the microphone M_i , α_i is the amplitude attenuation ratio and $n_i(t)$ is the noise signal picked up by the microphone M_i .

The Time Delay of Arrival (TDOA) is defined for a given microphone pair i and k as:

$$D_{ik} = \Delta_i - \Delta_k \quad (2.2)$$

The goal of the first step of the source location process is to determine D_{ij} for some subset of microphone pairs. The Fourier transform of the received signal in (2.1) can be expressed as:

$$m_i(t) \leftrightarrow M_i(\omega) = \alpha_i X_i(\omega) e^{-j\omega\Delta_i} + N_i(\omega) \quad (2.3)$$

It can be assumed that the average energy of the captured signals is significantly greater than that of the intrusive noise:

$$\alpha_i^2 \overline{|X(\omega)|^2} \gg \overline{|N_k(\omega)|^2} \text{ for all } i, k \quad (2.4)$$

In addition, the signal to noise ratio (SNR) of the captured signal is defined to be:

$$\text{SNR}_i(\omega) \triangleq 10 \log \left(\alpha_i^2 \frac{\overline{|X_i(\omega)|^2}}{\overline{|N_i(\omega)|^2}} \right) \quad (2.5)$$

If the condition in (2.4) is true, the SNR and the cross correlation of $m_i(t)$ and $m_k(t)$ defined as:

$$R_{ik}(\tau) = \int_{-\infty}^{\infty} m_i(t) m_k(t - \tau) dt \quad (2.6)$$

can be expected to be maximum, at $\tau = D_{ik}$. The frequency domain representation of (2.6) is:

$$R_{ik}(\tau) \leftrightarrow S_{ik}(\omega) = M_i(\omega) M_k^*(\omega) \quad (2.7)$$

where $*$ is the complex conjugate operation. Equation (2.7) can be expanded using (2.3) as follows:

$$\begin{aligned} S_{ik}(\omega) &= (\alpha_i X_i(\omega) e^{-j\omega\Delta_i} + N_i(\omega)) (\alpha_k X_k^*(\omega) e^{j\omega\Delta_k} + N_k^*(\omega)) \\ &= \alpha_i \alpha_k |X_i(\omega)|^2 e^{-j\omega(\Delta_i - \Delta_k)} + N_i(\omega) N_k^*(\omega) \\ &\quad + \alpha_i X_i(\omega) e^{-j\omega\Delta_i} N_k^*(\omega) + \alpha_k X_k^*(\omega) e^{j\omega\Delta_k} N_i(\omega) \end{aligned} \quad (2.8)$$

We can consider the last three terms of (2.8) as negligible compared to the first term based on the assumption in (2.4). Expression of (2.8) now reduces to:

$$S_{ik}(\omega) \approx \alpha_i \alpha_k |X_i(\omega)|^2 e^{-j\omega D_{ik}} \quad (2.9)$$

We can find D_{ik} by evaluating:

$$D_{ik} = \max_{\tau} R_{ik}(\tau) = \max_{\tau} \mathcal{F}^{-1}\{S_{ik}(\omega)\} \quad (2.10)$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transformation operator.

2.2.2 Cross-Power Spectrum Phase (CPSP)

Given no *a priori* statistics about the source signal and the interfering noise, it can be shown [19] that the optimal approach for estimating D_{ik} is to whiten the cross-correlation by normalizing it by its magnitude:

$$\frac{S_{ik}(\omega)}{\alpha_i \alpha_k |X(\omega)|^2} = e^{-j\omega D_{ik}} \leftrightarrow \delta(\tau - D_{ik}) \quad (2.11)$$

Because the denominator term is not known in (2.11), we may again apply the approximation in (2.4) and normalize it by the product of the magnitudes of the captured signals. The described function is defined as the cross-power spectrum phase (CPSP) function:

$$\text{CPSP}_{ik}(\omega) \triangleq \frac{M_i(\omega) M_k^*(\omega)}{|M_i(\omega)| |M_k(\omega)|} \quad (2.12)$$

And in the time domain:

$$\text{cpSP}_{ik}(\tau) \triangleq \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{m_i(t)\} \mathcal{F}\{m_k(t)\}^*}{|\mathcal{F}\{m_i(t)\}| |\mathcal{F}\{m_k(t)\}|} \right\} \quad (2.13)$$

We can see from (2.11) that the output of the CPSP function is delta-like with a peak at

$$\tau = D_{ik}.$$

The abovementioned analysis is derived based on analog signals and stationary sets of D_{ik} . In digital implementations, the $m_i(t)$'s are sampled and converted into discrete sequences $m_i(n)$. In addition, the D_{ik} 's are not stationary, in view of the fact that the sound source is apt to change location. Finite frames of processing are also required due to computational constraints. Hence, the typical windowing techniques are applied and the sampled signals are broken into analysis frames. After conversion into the discrete finite sequence domain, expression (2.13) becomes:

$$\text{cpSP}_{ik}(n) = \text{IDFT} \left\{ \frac{\text{DFT}\{m_i(n)\} \text{DFT}\{m_k(n)\}^*}{|\text{DFT}\{m_i(n)\}| |\text{DFT}\{m_k(n)\}|} \right\} \quad (2.14)$$

where the DFT is the Discrete Fourier Transform and IDFT is its inverse operation.

2.2.3 Modified CPSP

The whitening of the cross correlation spectrum is based on the assumption that the spectral behaviour of both signal and noise is uniform across the entire spectrum. More specifically, it is assumed that the approximation in (2.4) is equally valid for all ω . In practice, the SNR level varies with ω . In untreated enclosures, there are usually large amounts of acoustic noise below 200Hz due to line frequency related acoustic noises and Hoth noise [95] [Subsection 7.3.3]. It is therefore desirable to discard the portion of the CPSP which is below that level. Moreover given that the source component $x(t)$ provides the dominant amount of overall energy in $m_i(t)$, it can be expected that there

is a higher SNR at frequencies where the magnitude of $M_i(\omega)$ is greater. It is therefore advantageous to weight the portions of $S_{ik}(\omega)$ with greater magnitude, more heavily. An alternate expression to (2.12) is proposed that performs only a partial whitening of the spectrum:

$$\text{CPSP}_{ik}(\omega) \triangleq \frac{M_i(\omega)M_k^*(\omega)}{(|M_i(\omega)||M_k(\omega)|)^\rho} \quad 0 \leq \rho \leq 1 \quad (2.15)$$

A discrete equivalent for the modified CPSP is expressed as:

$$\text{cpSP}_{ik}(\mathbf{n}) = \text{IDFT} \left\{ \frac{\text{DFT}\{m_i(\mathbf{n})\}\text{DFT}\{m_k(\mathbf{n})\}^*}{(|\text{DFT}\{m_i(\mathbf{n})\}||\text{DFT}\{m_k(\mathbf{n})\}|)^\rho} \right\} \quad 0 \leq \rho \leq 1 \quad (2.16)$$

Setting ρ to zero generates non-normalized cross-correlation, while setting ρ to one produces (2.14). A good value for ρ may be determined experimentally, and varies with the characteristics of the room noise and the acoustical reflectivity of room walls. An optimal value for ρ was determined to be about 0.75 for several different enclosures with different characteristics of room noise and the acoustical reflectivity of walls [20].

A separate study shows choosing $\rho = 0$ reduces the TDOA anomalies (outliers) in male speakers while $\rho = 1$ still works better for female speakers [21]. Having removed anomalies, $\rho = 1$ has an overall reduced RMS error for both male and female speakers.

2.2.4 Source Location Finding Algorithm

The goal of the location finding algorithm is to determine the location of the sound source based on a selected set of TDOA's. Consider a sound source S with coordinates $s = \{x_s, y_s, z_s\}$ and a microphone pair M_1, M_2 with coordinates $m_1 = \{x_{m1}, y_{m1}, z_{m1}\}$ and

$\mathbf{m}_2 = \{x_{m2}, y_{m2}, z_{m2}\}$ as shown in Figure 2-5, it takes time t_1 for sound to propagate from S to M_1 and time t_2 to propagate from S to M_2 .

Figure 2-5 Source locus in 2D [19]

A given time t_i of propagation may be computed by the parametric equation:

$$t_i = \frac{d(\mathbf{m}_i, \mathbf{s})}{V_{sound}} = \frac{\sqrt{(x_{m_i} - x_s)^2 + (y_{m_i} - y_s)^2 + (z_{m_i} - z_s)^2}}{V_{sound}} \quad (2.17)$$

where $V_{sound} = c$ is the speed of sound, which is approximately 340 meters/second at room temperature and sea level atmospheric pressure.

The TDOA computed for the pair $\mathbf{m}_1, \mathbf{m}_2$ defines the difference $t_1 - t_2$. This difference also defines a hyperplane H for which the difference $t'_1 - t'_2$ is also equal to the TDOA estimate D_{12} . All of the points lying on this plane are potential locations of source S . The hyper-plane is a three dimensional hyperboloid defined by the following parametric equation:

$$\frac{d(\mathbf{p} - \mathbf{m}_1) - d(\mathbf{p} - \mathbf{m}_2)}{V_{sound}} = t_1 - t_2 = D_{12} \quad (2.18)$$

Where $p = \{x_p, y_p, z_p\}$ defines a point on the hyper-plane H . Figure 2-6 shows the hyperboloid of constraint formed by a known delay between two microphones in 3D. Note that a pre-held constraint only considers a one sided hyperboloid given that sound can only propagate in one direction.

Figure 2-6 Source locus in 3D, hyperboloid hyper plane [4]

A given set of TDOA's $\{D_{ik}\}$ has an associated set of hyperplanes $\{H_{ik}\}$. The location of the sound source must be a point p that lies on every H_{ik} , and satisfies the set of associated parametric equations:

$$\left\{ \begin{array}{l} \frac{d(\mathbf{p}-\mathbf{m1})-d(\mathbf{p}-\mathbf{m2})}{V_{\text{sound}}} = D_{12} \\ \vdots \\ \frac{d(\mathbf{p}-\mathbf{mi})-d(\mathbf{p}-\mathbf{mk})}{V_{\text{sound}}} = D_{ik} \\ \vdots \end{array} \right\} \quad (2.19)$$

Thus theoretically a set of three $\{D_{ik}\}$ uniquely specifies the coordinates of the source.

For sets of four or more $\{D_{ik}\}$, which we have over-determined sets of equations; a

solution may only exist in the least mean square (LMS) sense. In practice, however, the two bearing planes will seldom intersect to one point due to detection error caused by environment noise and reverberation. Approximate closed form solutions of (2.19) exist in a two-dimensional case. In a three-dimensional case, an approximate solution may be obtained when special restricted arrangements of the microphone pairs are used [22]. In this approximate solution the hyperboloid hyperplane can be approximated by a cone (Figure 2-7). This cone lies on hyperboloid asymptotes. If three-dimensional resolution is required for arbitrary microphone arrangements, a closed form solution does not exist, and numerical methods must be used.

Figure 2-7 Source locus in 3D cone approximation of hyperboloid [4]

The closed-form, analytic, noniterative location estimation solution for intersection of hyperbolic curves first presented in [23]. This solution is optimum and approximates the maximum-likelihood estimator and attains the Cramer-Rao lower bound near the small error region.

A closed-form analytic solution for the three dimensional localization (hyperboloid triangulation) in the near field has been presented recently [24]. This method uses a minimum of five microphones in three dimensions, and claims that in non-singular layouts of the microphones performs fast and accurate.

2.3 Advanced Developments in Sound Source Localization

This section mostly deals with new enhancements and modifications in the sound source localization (SSL). Here we present the recent developments in different categories. To be comprehensive, we covered almost all significant research in this topic to date.

2.3.1 General

Some of these approaches are topology based [1, 25]. Dynamic selection of microphone pairs, rather than fixed *a priori*, through some special parameters and criteria for a suitable range for microphone-pair separation and an ideal length for the TDOA vector, is the goal of research in [26].

Other research has focused on modifying the traditional two step procedure to release the need for explicit time-delay estimates [27]. Instead, the cross-correlation functions derived from various microphone pairs are simultaneously maximized over a set of potential delay combinations consist with the candidate source locations. The result is a procedure that combines the advantages offered by the phase transform (PHAT) weighting (or any reasonable cross-correlation-type function) and a more robust localization procedure without dramatically increasing the computational load.

Others add models of human vocal tract to the propagation model and claim for exact speaker localization instead of sound source localization [28].

2.3.2 Distributed Arrays

Some researchers have focused on using distributed arrays in SSL and defining some spatial related factors to the localization algorithm [29, 30, 31]. Distributed networks of microphones have also been used to create acoustic maps based on the classification of a

global coherence field or oriented global coherence field to identify position and orientation of a speaker [32]. The near-and far-field arrays of such a distributed network is utilized in the European Commission integrated project CHIL, "Computers in the Human Interaction Loop" project, to solve the problems of speaker localization and tracking, speech activity detection and distant-talking automatic speech recognition [33]. Another method uses acoustic signal energy measurements taken at individual sensors of an ad hoc wireless sensor network to estimate the locations of multiple acoustic sources [34]. A multi-resolution search algorithm and an expectation-maximization like iterative algorithm are proposed to expedite the computation of source locations to track military vehicles. Another setting uses sparse array of arbitrary placed sensors (several laptops/PDAs co-located in a room) [35]. Therefore any far-field assumptions are no longer valid in this situation and the performance of the localization algorithm is affected by uncertainties in sensor position and errors in A/D synchronization. The proposed source localization algorithm consists of two steps. In the first step, time differences of arrivals are estimated for the microphone pairs, and in the second step the maximum likelihood estimation for the source position is performed.

2.3.3 Multi Rate, Spectral, Non-free Field

Multi-rate acquisition of sound is also studied to create a high quality sound signal out of lower rate sampling signals utilizing TDOA information [36, 37].

Aside from free field microphone array locators, some researchers have focused on human ear based methods [38]. They use spherical phantoms to mimic the head diffraction of the sound.

Sturim [39, 40] proposed an algorithm for tracking multiple speakers using a linear microphone array. The input for his algorithm is a time delay based localization estimator. As noted there, localization measurements are valid only during periods of single source speech activity. During time intervals of multiple talkers no localization observations are provided.

In the case of speech signals, certain characteristics can assist in distinguishing between the sources. One such feature is the spectral signature of each speaker [5, 7, 10]. Spectral differences between speakers stem from the physiological fact that each speaker has an individual system of speech organs, which are similar, but not identical to those of other speakers. These spectral differences due to different vocal aspects may be used to distinguish between different speakers. Working in the frequency domain enables transforming the problem of direction estimation from a wideband, multi-source problem into a set of single-source, single-frequency problems.

The spherical analysis of a sound field can provide spatial information about the individual reflections in an enclosure, which can then be used to compute several spatial room acoustics measures and leads us to a better understanding regard the main sound directions and also reverberations. One form of the sound field analysis is by the plane wave decomposition of a sound field, where the sound field is decomposed into sound field components. The formulation of the plane-wave decomposition from the pressure distribution on a sphere was presented by Rafaely [41]. A spherical microphone array with 98 microphones, designed around a rigid sphere, analyzed and simulated to decompose the sound field of an anechoic chamber and an auditorium into plane waves [42].

2.3.4 Orientation Assisted

A new modification known as enhanced sound localization, offers a joint localization and orientation estimation of a directional sound source using distributed microphones [43, 44]. The joint orientation and localization estimates are results of explicitly modeling the various factors that affect the level of access to different spatial positions and orientations for a microphone in an acoustic environment. Three primary factors are accounted for, namely the source directivity, microphone directivity, and source-microphone distances. Later, a multi-dimensional search over all possible sound source scene reconstruction algorithms is presented in the context of an experiment with 24 microphones and a dynamic speech source. At a signal-to-noise ratio of 20 dB and with a reverberation time of approximately 0.1 s, accurate location estimates (with an error of 20 cm) and orientation estimates (with an average error of less than 10 deg) are obtained [45, 46].

Measurements of a talker in an anechoic chamber have shown significant anisotropy in radiation patterns. Orientation estimation and its positive effect on the localization accuracy have also been addressed in another research [47]. Using only acoustic energy data obtained from a large-aperture microphone array (448 microphones) the head bearing of a talker within a large focal area is determined via a beamforming approach that employs an optimization function and used to improve location-estimation.

Considering the non-omnidirectionality of the sound source, the global coherence field (GCF) or SRP-PHAT has been modified to a more informative map called Oriented GCF (OGCF) [48, 49]. Using OGCF has shown an improved localization performance with respect to GCF. It has also been proposed to integrate localization obtained as

maximum peak of GCF or OGCF with a classification step considering the whole GCF or OGCF maps.

The inverse of the sound source localization problem has been used for localizing a microphone array when the location of sound sources in the environment is known [50]. Using a particular spatial observability function, a maximum likelihood estimator for the correct position and orientation of the array is derived. This is used to localize and track a microphone array with a known and fixed geometrical structure.

2.3.5 Optimization Based

Classical acoustic source localization algorithms attempt to find the current location of the acoustic source using data collected at an array of sensors at the current time only and with known determined geometry. In the presence of strong multipath, these algorithms occasionally locate a multipath reflection rather than the true source location. A recently proposed method is a state-space approach using particle filtering [51]. This approach formulates a general framework for tracking a moving acoustic source using particle filters and claims to track a moving source accurately in a moderately reverberant room. Particle filtering is also used to estimate the source location through steered beamforming [52]. This scheme is especially attractive in speech enhancement applications, where the localization estimates are typically used to steer a beamformer at a later stage. Related research effectively employs particle filtering and particle swarm optimization in an integrated framework [53]. Source localization is viewed as a global minimization problem where the solution is searched by properly exploiting competition and cooperation among the individuals of a population.

Recently, the approaches that are based on search for local peaks of the steered response power are becoming popular, despite their known computational expense. It has been shown that computing the steered response power is more robust than the faster, two-stage, time difference of arrival methods. The problem is that steered response power space has many local maxima and thus computationally-intensive grid-search methods are used to find a global maximum. Stochastic region contraction has been proposed to speed up the search [54].

Based on the observation that the wavelengths of the sound from a speech source are comparable to the dimensions of the space being searched and that the source is broadband, another research group developed an efficient search algorithm [55, 56]. Significant speedups are achieved by using coarse-to-fine strategies in both space and frequency. A fast spherical array beamformer has been designed based on the above algorithm [57].

2.3.6 TDOA Enhancement

Time delay of arrival is the basic technique for numerous applications where there is a need to localize and track a radiating source. Concentration on increasing the accuracy of time delay estimation via different methods is another approach that indirectly enhances the performance of sound source localization systems. Specific research employs more sensors and takes advantage of their delay redundancy to improve the precision of the TDOA estimate between the first two sensors [58]. The approach is based on the multi-channel cross-correlation coefficient and is claimed to be more robust to noise and reverberation. Later, they re-formulated the approach on a basis of joint entropy and showed that for Gaussian signals maximizing multi-channel cross-correlation is

equivalent to minimizing the joint entropy [59]. However, with the generalization of the idea to non-Gaussian signals (e.g. speech), the joint entropy-based new TDE algorithm manifests a potential to outperform the multi-channel cross-correlation-based method. Here [60], the multi-channel cross-correlation is related to the well-known linear interpolation technique and a recursive algorithm is introduced so that it can be estimated and updated efficiently. Experiments confirm that the relative time-delay estimation accuracy increases with the number of sensors.

Gaussianity of the source signal, often used in blind source separation algorithms, can be applied to define an information-theoretical measure called mutual information to enhance estimated TDOA under reverberant conditions [61, 62]. Sound features corresponding to the excitation source of the speech production mechanism are claimed to be robust to noise and reverberation. One of these impulse-like excitations, the Hilbert envelope of the linear prediction residual of voiced speech, is extracted reliably from the speech signal and used for time delay estimation [63]. This method is claimed to perform better than the generalized cross-correlation approach.

Similar to spectrogram, two useful visualization tools for visualizing phasic behaviours of microphone arrays has been introduced to visually analyze the effect of reverberation on TDOA [64].

2.3.7 Pre-filtering and Clustering

Filtering and clustering of outliers (anomalies) either in TDOA stage or in the final localization stage is another area for SSL enhancement. Prefiltering is usually adopted to reduce the spurious peaks due to reflections. The secondary peaks of the generalized cross correlation can be crucial in order to correctly locate the sound source. An iterative

weighting procedure is introduced based on this rationale and peaks corresponding to the actual source position consistently weighted. The position estimate is then refined using an effective and fast clustering technique [65].

Room reverberation is typically the main obstacle for designing precise microphone-based source localization systems. Dereverberation approaches like cepstral prefiltering have been proposed, but they are computationally expensive and inadequate for real-time applications. A research group developed a statistical model for the room transfer function [66]. They applied the image method for simulation of the room transfer function and carried out the asymptotic error variance and the probability of an anomalous outlier TDE estimates. Another pre-filtering approach, based on the common acoustical pole modeling of the room transfer functions, is presented and compared with existing techniques in [67].

An approach based on a disturbed harmonics model of time delays in the frequency domain employs the well-known ROOT-MUSIC algorithm, after suitable pre-processing of the received signals [68, 69]. Final clustering of raw TDOA estimates gives candidate source positions.

2.4 Orthogonal Microphone Work

The idea of using three orthogonally placed sensors, which will be discussed in the next chapter, is new but there have been similar ideas to detect a narrowband wave with four orthogonally spaced sensors on a circle. The Adcock -Butler direction finder is the first of such work [70, 71]. It consists of four identical elements arranged at the four corners of a square. The angle of arrival is determined by processing the difference of signals

(dividing and taking arctangent of the result) from the opposite pairs of the elements [72,73].

In our work, contrary to the Adcock-Butler topology, a cell has three sensors, one in the centre and the other two in two orthogonal directions and it operates on wideband signals by ensemble averaging.

In another work [74], two closely separated microphones with variable delays have been used so that by changing the delays the resulting beam pattern changes. By appropriately combining three sets of orthogonal pairs with simple scalar weightings, a general differential microphone beam can be realized and directed to any angle in space. The system of [68] has a wide directivity pattern and therefore it is suitable for short distance directional microphone applications.

Unfortunately there is no unified condition for scientific comparison of the different methods. Different rooms have different acoustical conditions and different amounts of signal-to-noise ratios as well as different paths of reverberations. For this very reason, in each of the previous works, the new contribution is verified only by comparing it to the a starting method. Although an anechoic chamber is commonly used in various acoustics research work, it is not a realistic enclosure because of the lack of reverberations on it.

2.5 Summary

To date there has been no work on sound source localization, to the best of our knowledge, that employs a 3-microphone orthogonal cell element, or employs a hemispherical array of microphones with a microphone in the center, or uses a fully time domain computation to estimate a wideband sound source direction as proposed in this thesis.

In examining the previous literature we also demonstrate that previous approaches (e.g. TDOA which is the fastest among them) require greater computational load than the method in this thesis [Subsection 7.3.6].

In terms of accuracy, since there is no common experimental test-bed employed by all researchers it is difficult to definitively compare any absolute accuracy claims with our system accuracy.

3. Eye Array in Sound Source Localization

“There is always another way to look at the same problem”

Richard P. Feynman, 1918-1988.

In this chapter, we will discuss our novel speaker localization method and describe the general framework of the method.

3.1 A New Localization Method

To achieve the goal of this research we propose a new localization method and show how a source, such as a speaker, can be localized by a passive spherical sensor array through the use of simple multiple parallel delay and difference operations. This novel localization technique assumes no *a priori* knowledge of the other localization techniques. The key idea behind this source localization method is to use multiple (2 or more) element sensor array “cells” oriented in different directions on a hemisphere, calculate a closeness measure/function for each cell direction and finally estimate the source direction from the set of closeness functions.

This source localization method is designed to have the following attributes:

- Parallel computation: The source localization in 2π steradians [75, 76] is divided into N detecting solid angles; each area monitors a solid angle equal to $2\pi / N$ steradians.
- Simplicity: The detecting cell algorithm has to be as simple as possible.

- Real time performance: For the sake of speed and simplicity we bound our method to time domain, real time calculations and avoid any frequency domain and computationally expensive calculations.
- Nonlinearity: Since our objective is to develop a time domain algorithm, there would be no restriction to linear systems and we benefit from time domain nonlinear calculations.
- Three dimensional enclosed space: The target is a system for three dimensional sound source localization of one dominant sound source in an enclosed environment with a reasonable level of noise and reverberation.

3.2 Assumptions

We will make some assumptions, which are common in most sound source localization methods:

- The source is a point source distant from the array so that the wavefront is planar (far field assumption).
- The source signal is wideband; in other words, it spans many frequencies in a small time frame.
- The traveling medium is homogeneous and isotropic with constant speed of wave propagation.
- All of the sensors (microphones in our case) are omni-directional and they have zero mutual couplings.

3.3 Overview

We started with a common sense idea that by benefiting from symmetrical properties of geometrically structured arrays, one can significantly reduce the computational cost in an array processing system. A sphere, the set of all points in three dimensional space which have a constant distance from a fixed point in that space, is a perfect geometrically symmetric surface. A hemisphere, also maintains the equidistance property from the centre. Therefore spheres and hemispheres are excellent candidates for arrays in geometrically symmetrical localization strategies.

3.3.1 Hemisphere versus Sphere

Suppose we have an array of point sensors, lying on the surface of a sphere of radius r , which are uniformly spaced so that for every sensor on the sphere we can find another antipodal sensor on the other side of the sphere. If we compute the root mean square of the difference of each sensor signal with its delayed antipodal sensor signal for all antipodal sensor pairs, the pair of sensors that exactly coincides with the direction of arrival of the plane wave results in a minimum output. That is, the line between the antipodal pair which has the minimum output value among all of the other pairs is in the direction of arrival. In this configuration, the “look direction” can be steered to any direction in 4π steradians solid angle, because of the symmetrical shape of the sphere. Here, if the number of microphone pairs on the sphere is N , we need $2N$ microphones as well as N delay elements, N subtraction elements and N integration elements.

Now let us consider that instead of having a spherical shell of omni-directional microphones, we distribute half of the microphones on a hemisphere shell and put one additional microphone in the centre of the flat side of the hemisphere. Since each planar

wave passes the centre of the sphere, and has a constant distance to other points on the sphere (likewise antipodes on the sphere), half of the antipodal sensors can be replaced by one omni-directional sensor at the centre of the sphere. This central microphone acts as the “reference microphone” for all other microphones. Although this method lowers our span angle from 4π steradians to 2π steradians, it lowers the number of delay elements from N to one. It also prepares an obstacle free anterior space for posterior microphones, for the reason that all of the anterior microphones have been replaced by one microphone in the centre. This array structure also solves the commonly known problem of front-back misjudgement [82], compared to linear microphone arrays.

3.3.2 Method Synopsis

Postulate that we have a hollow hemispherical structure distributed with a number of omni-directional microphones on its shell plus one omni-directional microphone located at the sphere center. Our goal is to define a closeness function for each microphone on the shell which provides the maximum (or minimum) closeness function output value if the sound source coincides with the corresponding shell microphone direction, and monotonically lower (or higher) closeness function output values elsewhere. A shell microphone direction is any of the directions parallel to the line from the center (“reference microphone”) to the corresponding microphone on the shell. Obviously the sound source direction is not always in the direction of one of the shell microphones. Therefore we have to estimate the direction by interpolating the outputs of adjacent cells. This interpolation can be done with a minimum number of neighbouring cells. The minimum number of points that can span a solid angle is three and the resulting area created by them forms a spherical triangle [76].

If the closeness function, which is a measure of the vicinity of the source frontal planar wave norm and the shell microphone direction, form a linearly decreasing (or increasing) output value around the shell microphone direction, then the interpolation of the source direction from the neighbouring sensor cell output values would be reduced to a simple linear weighted average. In this regard, the shape of a closeness function output value versus angle is a narrow isosceles triangle (in 2D) or a narrow cone (in 3D).

Having the dominant neighbour directions (at least three) with their corresponding closeness values, as well the angular information of their direction (fixed by their physical location), one can easily estimate the sound source direction with some simple weighted average calculations.

Figure 3-3 illustrates a virtual representation of the coverage of a spherical triangle with three adjacent closeness functions.

Figure 3-1 A spherical triangle coverage by three closeness functions

3.4 Definitions

Here, some of the concepts which are regularly used in the course of this thesis are described and labelled. There are more elaborate descriptions for some of these concepts later in this thesis.

- **Shell microphone:** Any of the microphones distributed on the hemispherical shell.
- **Reference microphone:** The microphone which is located at the center of the full sphere.
- **Microphone direction:** Any direction parallel to the line from the center (reference microphone) to the corresponding microphone on the shell.
- **Cell:** The set of microphones containing at least the reference microphone and one or two of the shell microphones. In this case of a two-microphone cell, the direction of the second microphone of the cell is called the cell direction. We can also have three-microphone cells, in which there is another microphone in the cell so that its direction is orthogonal to the direction of the second microphone.
- **Deviation angle:** The angular difference between the direction of each cell and the direction of the sound source.
- **Closeness function:** A function which is universally defined for each microphone cell. The input to this function is the signal from all of the cell microphones and there is one output signal for each cell so that it reaches a maximum (or minimum) output value when the sound source coincides with the cell direction and monotonically lower (or higher) output values versus

deviation angle. Monotonicity guaranties that each closeness function uniquely evaluates the sound source direction.

- **Estimation function:** The input to this function is all of the closeness function output values plus their fixed bearings and the output is the final estimation of the sound source bearing in the form of azimuth and elevation angles.

As described so far, our estimation strategy is based on using a collection of microphones distributed on a hollow hemispherical structure. Triples or pairs of these microphones create detecting cells. Although we may set a special physical structure for each of the detecting cells, different closeness functions may be applied to them. Depending on the closeness function algorithm, our estimation function changes. During the course of this project, we started with a two-microphone and later a three-microphone closeness function based on subtraction of the reference signal from shell signals. Later we defined two more sophisticated closeness functions which yielded better results than the first idea. Here in this thesis we start with the description of the final closeness function and cell, followed by a discussion on the previous systems and comparison of them.

3.5 Two-Microphone Cell

Consider two omni-directional microphones, which are separated by a displacement r in a placement in order that m_0 is located in the center and m_i is located somewhere on the shell of a hemisphere (Figure 3-4). The sound source is also located far from microphones to comply with far field assumption (Sections 3.2). Obviously we can shift

the time origin of the sound source to the center of hemisphere, without loss of generality. Therefore:

$$S_0(t) = S(t) + n_0(t) \quad (3.1)$$

and

$$S_i(t) = S(t - \tau_i) + n_i(t) \quad (3.2)$$

Figure 3-2. Representations in spherical coordinates

where i is the microphone index, τ_i is the time it takes for the sound to propagate from the center microphone (m_0) to the i^{th} microphone (m_i), and $n_i(t)$ is the sum of the noise and the reverberation signals acquired by m_i .

Since $S(t)$ is a natural sound signal and consequently differentiable, the Taylor series [77] expansion of the signal at the i^{th} microphone about the center microphone signal is convergent and can be written as:

$$S(t - \tau_i) = S(t) - \tau_i \dot{S}(t) + \frac{1}{2!} \tau_i^2 \ddot{S}(t) - \dots \quad (3.3)$$

Assume $S(t)$ is bandlimited and τ_i is small, therefore we can neglect the 2nd and the higher order terms:

$$\tau_i \dot{S}(t) \cong S(t) - S(t - \tau_i) \quad (3.4)$$

Considering (3.1), (3.2) and (3.4) we have:

$$\tau_i \dot{S}(t) \cong [S_0(t) - S_i(t)] - [n_0(t) - n_i(t)] \quad (3.5)$$

Here $\dot{S}(t)$ is the time domain derivative of the source signal ($\dot{S}(t) = \partial S(t) / \partial t$). We can calculate $\dot{S}(t)$ numerically. There are various numerical methods for time derivatives. Some of them are causal (depend on present and past), some are anticausal (depend on present and future) and some are noncausal (depend on past, present and future). The simplest causal approximation for the first derivative of $S(t)$, $\dot{S}(t)$ is the Backward Difference [77].

$$\dot{S}(t) \cong \frac{1}{T} [S(t) - S(t - T)] \quad (3.6)$$

Here T is the derivation time. The reference microphone signal $S_0(t)$, best represents the source signal $S(t)$. Therefore, considering (3.1) and (3.6) we have:

$$\dot{S}(t) \cong \frac{1}{T} [S_0(t) - S_0(t - T)] + \frac{1}{T} [n_0(t) - n_0(t - T)] \quad (3.7)$$

From (3.5) and (3.7):

$$\frac{1}{T} [S_0(t) - S_0(t - T)] \cong \frac{1}{\tau_i} [S_0(t) - S_i(t)] + n(t) \quad (3.8)$$

and

$$n(t) = -\frac{1}{\tau_i} [n_0(t) - n_i(t)] - \frac{1}{T} [n_0(t) - n_0(t - T)] \quad (3.9)$$

Note that the left side of (3.8) is the time domain gradient (measurement) and the right side is the spatial domain gradient (observation) multiplied by inverse of the time traveling delay (parameter) of the sound source signal plus noise. This equation is an affine transformation of parameter $(\frac{1}{\tau_i})$, so we can define a Euclidian objective function as:

$$J(\tau_i) = \left\| \frac{1}{T} [S_0(t) - S_0(t-T)] - \frac{1}{\tau_i} [S_0(t) - S_i(t)] \right\|^2 \quad (3.10)$$

This objective function can be minimized as a standard least mean square (LMS) problem. Hence τ_i can be obtained by [78,79]:

$$\frac{\tau_i}{T} = \frac{\overline{\langle [S_0(t) - S_i(t)] \cdot [S_0(t) - S_0(t-T)] \rangle}}{\overline{\langle S_0(t) - S_0(t-T) \rangle^2}} \quad (3.11)$$

The term $\frac{\tau_i}{T}$ was chosen as our first closeness function candidate. It has a maximum value, where the sound source direction coincides with the i^{th} microphone direction. As well it monotonically decreases with angular elevation of the sound source from the i^{th} microphone direction, in the close vicinity of the microphone direction. In fact, we can easily observe:

$$\tau_i = T_r \cos(\Theta) \quad (3.12)$$

where Θ is the spatial angle between the sound source direction and the i^{th} microphone direction and T_r is the time it takes for the sound to travel the hemisphere radius r :

$$T_r = \frac{r}{c} \quad (3.13)$$

where c is the sound speed. Therefore $\frac{\tau_i}{T}$ has a cosine shape drop off:

$$\frac{\tau_i}{T} = \frac{T_r}{T} \cos(\Theta) \quad (3.14)$$

which is nonlinear as well as very flat in the close vicinity of the i^{th} microphone direction ($\Theta \cong 0$).

Obviously this is not a desired characteristic intended for closeness functions, based on our initial goals declared in Section 3.3.

Figure 3-5 shows a visual representation of a hemisphere with two-microphone cosine shaped closeness functions.

Figure 3-3 Visual representation of two-microphone cosine shaped closeness functions

3.6 Spherical Travel Time Delay

To understand all aspects of our hemispherical array, it is beneficial to discuss the nature of the delay between the center and the shell of the sphere in detail. This provides us with some clues as to how to develop the structure of detector microphone cells.

Consider a sphere with radius r as depicted in Figure 3-2. Each shell direction Θ can be identified with azimuth and elevation angles of θ and φ respectively. The microphone m_i on the sphere can be identified with the position vector Θ_i . Also, the source direction vector Θ_s , is the normal to planar waves emitting from source S .

$$\Theta_i\{\theta_i, \varphi_i\} = [\sin \varphi_i \cos \theta_i \quad \sin \varphi_i \sin \theta_i \quad \cos \varphi_i]^T \quad (3.15)$$

$$\Theta_s\{\theta_s, \varphi_s\} = [\sin \varphi_s \cos \theta_s \quad \sin \varphi_s \sin \theta_s \quad \cos \varphi_s]^T \quad (3.16)$$

The travel time of a planar wave from the centre of the hemisphere with radius r to the microphone i on the shell is:

$$\tau\{\theta_s, \varphi_s, \theta_i, \varphi_i, r, c\} = \frac{r}{c} \Theta_i \cdot \Theta_s \quad (3.17)$$

which $\cdot$ is the inner product operator. This source direction dependent geometrical travel time (delay) is the main physical aspect on which our method is based and is shown in Figure 3-4 (left) through simulation of (3.17). Obviously, this delay has a cosine shape, viewing at the sound source axis.

3.6.1 Derivative Sampling

Although based on the Nyquist theorem the minimum sampling frequency for a signal with bandwidth of ω is 2ω , in time domain applications we need more samples from the bandlimited signal to extract the time domain information. This can either be done with oversampling or upsampling of a Nyquist sampled signal.

In digital simulation or digital implementations, the signals are sampled at a limited sampling frequency (f_s). Therefore, the maximum detectable time delay between two signals is $T_s = 1/f_s$. If we redraw the delay function based on this fact with a sampling

frequency of e.g. $f_s = 10\text{Khz}$, the quantized delay between the reference microphone and each microphone on the spherical shell is as illustrated in Figure 3-4 (right).

Figure 3-4 Continuous (left) and quantized (right) travel time from the reference microphone to a spherical shell

As illustrated on Figure 3-4, due to the cosine profile of the delay element in the two-microphone cell, we have the least angle detection resolution in and around the source direction, the direction in which we need the highest resolution. On the contrary, the highest angular resolution is in and around the points that are orthogonal to the sound source direction. There are three distinct methods to solve the resolution loss caused from cosine behaviour of the delay:

- Increase the radius r of the hemisphere
- Increase the sampling frequency of signals
- Implementation in the analog domain

The first solution is against the broad tendency toward reducing the size of the array. The second solution is preferred in digital implementations and can be performed by either increasing the sampling frequency or upsampling of the Nyquist sampled signals. Figure 3-4 (right) shows that the quantized delay resolution is higher for the points which are 90° away from the source direction. This implies that to increase the resolvability of a cell, one can utilize the microphones that are orthogonal to the source axis. A cell modification and enhancement will be described later in Section 3.7.

3.6.2 Derivation Time

The approximation in (3.4) is valid regardless of the derivation time T . Theoretically the approximation in (3.8) tends to equality, if in some way T becomes equal to τ_i . Since τ_i is unknown and varying with deviation angle, we can select a value for T corresponding to the direction that we need the highest accuracy. Where the sound source direction coincides with the microphone direction ($\Theta = 0$), we need the maximum estimation accuracy. At this situation τ_i is equal to T_r and therefore the best choice for T is T_r . Upon this when $\Theta = 0$, the time difference in (3.8) equals the spatial difference.

3.6.3 Processing Time Frame and Noise

For ergodic processes, time averages converges to statistical averages by increasing the time frame of ensemble averaging [80].

The time interval, the time length of each data block for a desired output, for most of the signal processing systems is defined by the noise and interference of the environment. Improvement in the characteristics of signal processing systems is

fundamentally dependent on the processing time. If the process time is high, the effect on energy characteristics, resolution, and noise immunity is high [81].

There are two main classes of noise and interference: additive and multiplicative. The effect of the addition of noise and interference to the signal generates an appearance of false information in the case of additive noise, especially if information is embedded in amplitude. Stochastic distortions of the signal are attributed to unforeseen changes in instantaneous values of the phase and amplitude as a function of time and can be considered as additive noise. Multiplication of signals containing additive noise generates multiplicative noise. The impact of multiplicative noise on signal parameters and qualitative characteristics of signal processing algorithms essentially depends on the relationship between the time interval (frame length), within the limits of which the signal is processed and analyzed plus the rate of change of the signal phase and amplitude as a function of time (bandwidth). If the time interval of any type of correlation of the signal with multiplicative noise is higher than the duration of information coherency of the signal (might be lower than signal length), then the effects of multiplicative noise on the correlated output are low.

Consequently there is a tendency to increase the time frame of averaging in (3.11) (duration of the moving average) to achieve better statistical accuracy and noise immunity. However, this in return increases the response lag of the system. Since in most sound source localization applications the sound source movement is not fast, there is plenty of time for ensemble averaging in those applications. In frequency domain methods (FFT based), a linear increase in frame length tends to an exponential increase in computational complexity, while in this method, since all of the calculations are performed directly in the time domain, a linear increase in frame length increases the

computational costs linearly. In our digital implementation, since we have used recursive methods for the averaging parts (to be discussed in Subsection 7.3.5), a linear increase in frame length results in almost no added computational cost, but some extra memory and longer initial settling time or response delay.

3.7 Three-microphone Cell

We found that the resolution of the two-microphone difference cell is not satisfactory around its detection axis. This in fact is an inherent characteristic of any endfire array. On the contrary the direction of arrival resolution for any uniform linear array is maximum at the broadside [82].

To resolve the problem of slow and nonlinear drop off, we present a modification in our initial two-microphone cell configuration by adding a third microphone. This third microphone has to be placed so that its direction is orthogonal to the direction of the cell microphone.

From now on we assume that our microphones have a special distribution on the shell so that for each microphone m_i on the shell, there is at least another microphone on the shell (m_j) so that their directions are orthogonal to each other $m_j = m_i^\perp$. Notice that if microphone pair direction $\overline{(m_0, m_i)}$ is along the sound source direction $\overline{(s, m_0)}$ (endfire situation), then the microphone pair direction $\overline{(m_0, m_i^\perp)}$ would be orthogonal to the sound source direction $\overline{(s, m_0)}$ (broadside situation). Similar to m_i , $m_i^\perp$ is one of the shell microphones therefore equation (3.11) not only holds for pairs $[S_0(t), S_i(t)]$, but also holds for $[S_0(t), S_i^\perp(t)]$. Thus:

$$\frac{\tau_i^\perp}{T} = \frac{\overline{\langle [S_0(t) - S_i^\perp(t)] \cdot [S_0(t) - S_0(t-T)] \rangle}}{\overline{\langle S_0(t) - S_0(t-T) \rangle}^2} \quad (3.18)$$

Let us append a spherical coordinate to the three-microphone plane, as depicted in Figure 3-4. The angle θ_i is the elevation angle of the sound source direction from the i^{th} direction and ϕ_i is the azimuth angle or deviation of the sound source direction from the i^{th} three-microphone cell plane. Therefore:

$$\frac{\tau_i}{T_r} = \cos \theta_i \sin \phi_i \quad (3.19)$$

$$\frac{\tau_i^\perp}{T_r} = \sin \theta_i \sin \phi_i \quad (3.20)$$

By dividing (3.19) by (3.20), we come across a more useful function:

$$\frac{\tau_i}{\tau_i^\perp} = \cot g(\theta_i) \quad (3.21)$$

Likewise dividing (3.11) by (3.18) leads to:

$$\frac{\tau_i}{\tau_i^\perp} = \frac{\overline{\langle [S_0(t) - S_i(t)] \cdot [S_0(t) - S_0(t-T_r)] \rangle}}{\overline{\langle [S_0(t) - S_i^\perp(t)] \cdot [S_0(t) - S_0(t-T_r)] \rangle}} \quad (3.22)$$

As we observe from (3.21), $\frac{\tau_i}{\tau_i^\perp}$ behaves as cotangent of the elevation angle. We know that at smaller angles, the cotangent function is approximately proportional to the inverse of the angle:

$$\cot g(\theta_i) = \frac{1}{\theta_i} - \frac{\theta_i}{3} - \frac{\theta_i^3}{45} - \dots \cong \frac{1}{\theta_i} \quad \forall \theta_i \text{ small} \quad (3.23)$$

Therefore if we choose $\frac{\tau_i}{\tau_i^\perp}$ (or $\frac{\tau_i^\perp}{\tau_i}$) as our closeness (or distance) function, we can

interpolate the sound source direction by a simple linear weighted average of the

spherical directions of the cells having the highest (or lowest) outputs. Hence this function defined on three-microphone cell, roughly acts in accordance with the goals that we declared in Section 3.3.

3.8 Array Topology

To develop the aforementioned three-microphone cells, we need a special microphone distribution on the hemisphere so that for each microphone direction we can find at least another microphone with an orthogonal direction. As well, we have to find the best pattern on which N points on a hemisphere are located so that they subdivide the hemisphere into nearly equal solid angle portions.

One topologically feasible choice is a hemi-polyhedron pattern (geodesic dome). There are five Platonic solids, which are tetrahedron, cube, octahedron, dodecahedron, and icosahedron. Higher orders of tessellation [83] of these Platonic solids or other polyhedrons produce closer approximations to a sphere (or hemisphere). We decided to utilize a geodesic structure; since its mechanical structure is modular and its node distribution is nearly uniform (Figure 3-5, Figure 7-3, Figure 7-5, Figure 7-6).

Among geodesic structures we found the two-frequency tessellation of Icosahedron, which can be easily subdivided into two parts, as the best choice. This topology, subdivides the frontal view solid angle (2π steradians) into 40 spherical triangles. It contains 10 equilateral triangles and 30 isosceles triangles. Groups of five of these isosceles triangles are used to replace each pentagon in the topology. The isosceles spherical triangles have angular length of 31.7 degrees between the equal sides and 36 degrees between the opposite side while equilaterals have 36 degrees angular length (Figure7-3). This creates a semi uniform spatial sampling on the hemisphere.

In this topology, known also as Buckminster Fuller's geodesic dome [83], every vertex has at least four other orthogonal vertices in diverse locations (Table 7-1). Having 26 vertices on the shell and one center point, we need 27 microphones to cover the whole 2π steradians frontal part of the array. That is a reasonable number of channels, considering simple time domain calculations required.

Figure 3-5 shows three different views of this microphone array arrangement. In Chapter 7 we will describe and illustrate this topology in more detail.

Figure 3-5 Two-frequency Icosahedral geodesic hemisphere microphone arrays

3.9 Algorithm

This section describes the eye array algorithm formulation utilizing multiplicative closeness functions (MCF). The digital version of the algorithm is discussed in Appendix A. The reader can also refer to Figure 7-16 and Figure 7-17 in Subsection 7.3.5 for the implemented software. As we discussed earlier all our effort was to simplify the final algorithm so that it would be implementable on an integrated circuit.

Note that, we first implemented a simple switch algorithm based on the RMS value of the central microphone signal to remove the time regions that corresponds to the weak

or no sound signal frames on all channels. This prevents the SSL algorithm to be applied on very low SNR portions of the sound signal.

3.9.1 Formulation

Note that the numerator and denominator in (3.22) are similar calculations. In both, their left hand term is the time derivative of the reference microphone signal and the right hand term is the spatial difference calculated over the i^{th} direction (numerator) and its diagonal counterpart $i^\perp$ (denominator). Notice that $i^\perp$ is not only the diagonal of i but also one of the main directions itself. Therefore we simply compute:

$$F_i(t) = \overline{\langle [S_0(t) - S_i(t)] \cdot [S_0(t) - S_0(t - T_r)] \rangle} \quad (3.24)$$

once for all directions ($\forall i = 1, 26$).

Later by dividing each $F_i(t)$ by its diagonal counterpart $F_{i^\perp}(t)$, we achieve all desired closeness functions.

In our selected geometry, every vertex has at least four other orthogonal vertices in various locations (Table 7-1). This is a surprising attribute of this topology, which is rare amongst other platonic geodesic hemispheres. These diverse orthogonal vertices are used to enhance averaging of the closeness function estimates. Therefore we took the root mean square of all closeness functions for each direction.

Later we sort all averaged closeness function estimates and choose the first k closeness functions with the highest values among them. Theoretically, the minimum suitable number for k is 3, by which we define a spherical triangle. But in practice, in the case of $k=3$, we encountered abrupt changes when the sound source direction passes the border from one spherical triangle to the other. This can be avoided by incorporating

more nodes/cells into the final estimation process (increasing k). Practically, in the case of $k = 5$ we achieved a smooth as well as accurate result. Thus, in addition to the first three maximum closeness functions, we consider two additional nodes related to closeness functions with the fourth and fifth rank. These two extra nodes are typically neighbors of the main spherical triangle with the highest closeness function values.

Having the k directions with highest averaged closeness function in hand, the last step is a simple weighted averaging of the k corresponding node azimuth and elevation angles to calculate the estimated azimuth and elevation angle of the sound source direction.

$$\hat{\phi}_s(t) = \frac{\sum_{i^*} \overline{F_i}(t) \phi_i}{\sum_{i^*} \overline{F_i}(t)} \quad (3.25)$$

$$\hat{\theta}_s(t) = \frac{\sum_{i^*} \overline{F_i}(t) \theta_i}{\sum_{i^*} \overline{F_i}(t)} \quad (3.26)$$

where i^* denotes the k selected directions with the maximum closeness function values.

In addition ϕ_i and θ_i are fixed topology dependent azimuth and elevation angles of the

shell microphones [Table 7-1]. The pair $[\hat{\phi}_s(t), \hat{\theta}_s(t)]$ denotes the final estimated azimuth and elevation angles of the sound source direction.

The digital formulation, flowchart and details of the multiplicative closeness function based algorithm are discussed in Appendix A. Moreover the rationale behind the name eye array is discussed in Appendix C.

4. Alternative Closeness Functions

This chapter discusses two additional types of closeness functions based on a pinhole cell and a lens cell. These closeness functions were the first ideas that were developed in this research to approximate the closeness of a sound source direction to a specified cell direction. These methods are based on similarity measures between pairs of signals. We attempt to discuss them based on our initial geometrical description. Note that although these two closeness functions have low computation requirements, the first closeness function that we discussed in the previous chapter outperforms these two closeness functions in accuracy as will be explained later.

4.1 Difference Closeness Function (DCF)

The idea of difference based closeness functions started from the simple fact that, in an endfire setting if we subtract a delayed version of the former microphone signal from the latter microphone signal, the outcome should be theoretically zero while the sound source is exactly in line with the two-microphone direction. Heuristically one expects the zero (or small) RMS in output right in that direction and a monotonically increasing RMS output versus source angular deviation.

4.1.1 Difference Two-Microphone (Pinhole) CF

Consider two omni-directional microphones m_0 and m_i which are separated by distance r . For the plane wave $S(t)$ with spectrum $S(\omega)$ (Figure 4-1):

Figure 4-1 Two-microphone difference configuration

The point source signal is considered to be $S(t)$ at the point that it reaches the anterior microphone, then:

$$S_0(t) = S(t) + n_0(t) \quad (4.1)$$

The signal at the posterior microphone is a delayed version of the anterior microphone.

$$S_i(t) = S(t - \tau_i) + n_i(t) \quad (4.2)$$

The delay is a cosine function depending on the wave front angle of arrival Θ , the distance between the two microphones r , and the speed of the sound in the medium c :

$$\tau_i = \frac{r}{c} \cos(\theta) = T_r \cos(\theta) \quad (4.3)$$

We will not consider any amplitude loss due to wave travel, based on the far field source assumption. The amplitude difference between the delayed signal of the anterior microphone and the posterior microphone signal is:

$$y_i(t) = S_0(t - T_r) - S_i(t) \quad (4.4)$$

Considering (4.1), (4.2) and (4.4):

$$y_i(t) = [S(t - T_r) - S(t - \tau_i)] + [n_0(t - T_r) - n_i(t)] \quad (4.5)$$

By choosing a quadratic cost function to resemble the noise power:

$$J(y_i) = \|y_i(t) - [S(t - T_r) - S(t - \tau_i)]\|^2 = \|n_0(t - T_r) - n_i(t)\|^2 \quad (4.6)$$

The optimal maximum likelihood (ML) solution that minimizes J (Euclidian distance and noise power) is the sample mean or average [79]:

$$\hat{y}_i(t) = \arg \min [J(y_i)] = \overline{[S(t - T_r) - S(t - \tau_i)]} \quad (4.7)$$

To understand more about the two-microphone difference function behaviour (4.4), let's observe its frequency response. From (4.4):

$$Y(\omega, \theta) = S(\omega) [e^{-j\omega T} - e^{-j\omega T_r}] \quad (4.8)$$

$$H(\omega, \theta) = \frac{Y(\omega, \theta)}{S(\omega)} = 2j \sin \left[\frac{\omega T_r}{2} (\cos \theta - 1) \right] e^{\frac{-j\omega T_r}{2} (1 + \cos \theta)} \quad (4.9)$$

$$|H_i(\omega, \theta)| = \left| 2 \sin \left[\frac{\omega T_r}{2} (\cos \theta - 1) \right] \right| \quad (4.10)$$

Figure 4-2 displays (4.10), a 3D surface view along with a contour plot of the amplitude frequency response of (4.4) with respect to the angle of arrival θ in degrees and frequency f in hertz (Hz) ($T_r = 1. \text{ ms}$).

Figure 4-2 Frequency versus angle response of difference element

As illustrated, there is a cosine shape valley at and around $\theta \approx 0$ for all frequencies, but there are some other sharp zero crossings at other angles and different frequencies. Since we need a direction detector element for a broadband signal, we have to find a way to overcome the dilemma of occurrence of other zeroes. As illustrated, the zeroes at a particular angle are related to a small number of frequencies, which corresponds to the wavelength of the “seen path” at that angle and its higher harmonics. One way to overcome this predicament is utilizing a time averaging at the output. This time average acts as a low pass filter and smooths the response for broadband signals. Since it is assumed that the signal $S(t)$ covers all frequencies in a reasonable time frame, the time averaging operation removes all intermittent frequency zeros at “off-axis” angles for those frequencies. Therefore the only angle at which an optimum (minimum) occurs (which theoretically is zero) is the two-microphone axis (cell axis) $\theta = 0$.

This heuristic solution not only resolves the problem of extra zero crossings but also is the optimum for reducing the noise power as contemplated in (4.7).

If we assume that $S(t)$ uniformly covers all frequencies (white noise) at the duration of integration, the output response of the difference element with respect to sound source direction of arrival is as shown in Figure 4-3.

Figure 4-3 Angular response of integrated difference element

Obviously this type of output can be considered as a two-microphone detector cell. This output can be used as a measure of closeness of the source plane wave and difference cell in close proximity to the cell. But as illustrated the cosine dominated shape prevents us from having high resolving ability around the cell direction.

4.1.2 Difference Three-Microphone (Lens) CF

Similar to the preceding closeness function, a difference based lens cell utilizes another omni-directional microphone, orthogonal to the detection axis. Figure 4-4 shows the block diagram of this new three-microphone cell. The microphone m_0 is the reference

microphone in the center. The microphone m_i is the shell microphone as in the two-microphone cell. The microphone $m_{i\perp}$ is the orthogonal microphone.

Figure 4-4 Lens microphone cell

In this configuration, in addition to the previous cell, we will consider a term, which is the difference of the reference microphone m_0 and lateral microphone m_i . As we have seen earlier in section 4.1.1, the output of a two-microphone closeness function approximates a cosine function in vicinity of $\theta \approx 0$. This was indeed the main reason for low detection resolution around the detection axis. The idea of adding another microphone in the detection cell is to add a sine related term to the detection cell so that by dividing those two terms we could reach a tangent (cotangent) related term which is sharper in the vicinity of the main axis and is also linear (inversely linear) versus angle. Therefore, we expect higher resolution as well as a linear response with this modification. By choosing the endfire and broadside terms similar to:

$$y_{Def}(t) = \overline{S_0(t - T_r) - S_i(t)} \quad (4.11)$$

$$y_{Dbs}(t) = \overline{S_0(t) - S_{i\perp}(t)} \quad (4.12)$$

The difference lens closeness function can be defined as:

$$\bar{y}_{DCF}(t) = \frac{y_{Def}(t)}{y_{Dbs}(t)} = \frac{\overline{S_0(t-T_r) - S_i(t)}}{\overline{S_0(t) - S_{i_1}(t)}} \quad (4.13)$$

We can use the optimally equivalent terms, similar to (4.7), for the numerator and denominator of (4.13):

$$\hat{y}_{DCF}(t) = \frac{\overline{S(t-T_r) - S(t-\tau_i)}}{\overline{S(t) - S(t-\tau_{i_1})}} \quad (4.14)$$

4.1.2.1 Pseudo Transfer Function

Although the lens cell is a nonlinear system, we attempt to understand its behaviour based on the underlying mathematics. If we evaluate one instance of the time averaged terms in (4.14), in order to ignore the statistical average and merely consider $\tau_i = T_r \cos(\theta_i)$ and $\tau_{i_1} = T_r \sin(\theta_i)$:

$$y_{DCF}(t) = \frac{S(t-T_r) - S(t-T_r \cos(\theta))}{S(t) - S(t-T_r \sin(\theta))} \quad (4.15)$$

Equation (4.15) shows a division type nonlinearity. Both the numerator and the denominator are linear. In nonlinear systems superposition does not hold. Therefore, a frequency analysis has to be considered with caution. In linear systems, a transfer function is defined as the Fourier transform of the system impulse response. Equivalently, it is the response of the system to a single complex frequency phasor $Ae^{j\omega t}$ in the time domain. Here, similarly, we find the time domain response of the nonlinear system of (4.15) and try to characterize its response as a transfer function, at least in a specified angular region.

Therefore if:

$$S(t) = Ae^{j\omega t} \quad (4.16)$$

Then:

$$y_{DCF}(t, \omega, \theta) = \frac{Ae^{j\omega(t-T_r)} - Ae^{j\omega[t-T_r, \cos(\theta)]}}{Ae^{j\omega t} - Ae^{j\omega[t-T_r, \sin(\theta)]}} = \frac{Ae^{j\omega t} \cdot e^{-j\omega T_r} (1 - e^{-j\omega T_r [\cos(\theta) - 1]})}{Ae^{j\omega t} (1 - e^{-j\omega T_r \sin(\theta)})} \quad (4.17)$$

$$y_{DCF}(t, \omega, \theta) = \left(-\frac{\sin[\frac{\omega T_r}{2} [1 - \cos(\theta)]]}{\sin[\frac{\omega T_r}{2} \sin(\theta)]} \right) e^{\frac{j\omega T_r}{2} [1 + \cos(\theta) - \sin(\theta)]} \quad (4.18)$$

Therefore:

$$|y_{DCF}(t, \omega, \theta)|_{S=e^{j\omega t}} = \left| \frac{\sin[\frac{\omega T_r}{2} [1 - \cos(\theta)]]}{\sin[\frac{\omega T_r}{2} \sin(\theta)]} \right| \quad (4.19)$$

As demonstrated in (4.19), the numerator is the magnitude of the transfer function of the endfire term while the denominator is the magnitude of the transfer function of the broadside term. The fact that this response is not a function of time explains that the output of (4.15) to a tone is a DC signal. In general, this definitely does not hold for the response of the system to multi-frequency inputs, due to the non-validity of superposition. The 3D surface view and 2D contour plot representation of (4.19), the pseudo frequency response of (4.15), is shown in Figure 4-5.

Figure 4-5 Pseudo frequency versus angle response of difference lens CF

As depicted in Figure 4-5, this response has a linear region around $\theta \approx 0$, which is wide for low frequencies and narrow for high frequencies.

For small θ and low frequency signals, we can make the following approximations:

$$\sin\left[\frac{\omega T_r}{2}(1 - \cos(\theta))\right] \cong \frac{\omega T_r}{2}(1 - \cos(\theta)) \quad (4.20)$$

$$\sin\left(\frac{\omega T_r}{2} \sin(\theta)\right) \cong \frac{\omega T_r}{2} \sin(\theta) \quad (4.21)$$

Therefore (4.19) simplifies to:

$$\left| y(t, \omega, \theta) \right|_S = e^{j\omega t} \cong \left| \frac{\frac{\omega T_r}{2}[1 - \cos(\theta)]}{\frac{\omega T_r}{2} \sin(\theta)} \right| \cong \left| \frac{1 - \cos(\theta)}{\sin(\theta)} \right| = \tan\left(\frac{\theta}{2}\right) \approx \frac{\theta}{2} \quad (4.22)$$

For small θ and high frequency signals, we can use these approximations:

$$1 - \cos(\theta) \cong \frac{\theta^2}{2} \quad \sin(\theta) \cong \theta \quad (4.23)$$

Yet again (4.19) simplifies to:

$$|y(t, \omega, \theta)|_{S=e^{j\omega t}} \cong \left| \frac{\sin(\frac{\omega T \theta^2}{4})}{\sin(\frac{\omega T \theta}{2})} \right| \approx \frac{\theta}{2} \quad (4.24)$$

Note that the range of approximation validity at higher frequencies is lower than that at lower frequencies, and this is the main reason for the horn-like shape of the linearity region in Figure 4-5. Therefore, there is a region around $\theta \approx 0$ in which the output response is linear versus θ . For larger θ 's, we cannot expect (4.24) to be valid in multi frequency cases. Since nonlinear systems often have spectral spreading effects on their input signals, the analytic analysis of that region is not mathematically straightforward.

Yet again if we assume that $S(t)$ sweeps all frequencies uniformly at the specified duration of integration time, the output response of a difference type lens closeness function with respect to deviation angle from the cell axis is as follows:

Figure 4-6 Angular response of integrated lens cell

A similar outcome is accessible in time domain considering two approximations.

Keeping two terms of the $S(t - \tau)$ Taylor series expansion in equation (3.3), results in:

$$S(t - \tau) \approx S(t) - \tau \dot{S}(t) \quad (4.25)$$

Considering (4.14) and (4.25), one can rewrite the three microphone difference closeness function:

$$\hat{y}_{DCF}(t) \approx \frac{\overline{S(t - T_r) - S(t) + \tau_i \dot{S}(t)}}{\overline{\tau_{i_\perp} \dot{S}(t)}} = \frac{\overline{\tau_i \dot{S}(t) - T_r \left(\frac{S(t) - S(t - T_r)}{T_r} \right)}}{\overline{\tau_{i_\perp} \dot{S}(t)}} \quad (4.26)$$

The term in parenthesis is the Backward Difference approximation of $\dot{S}(t) = \partial S(t) / \partial t$.

Using this approximation:

$$\hat{y}_{DCF}(t) \approx - \frac{\overline{\left(1 - \tau_i / T_r \right) \dot{S}(t)}}{\overline{\left(\tau_{i_\perp} / T_r \right) \dot{S}(t)}} = \frac{1 - \cos(\theta)}{\sin(\theta)} = \tan\left(\frac{\theta}{2}\right) \approx \frac{\theta}{2} \quad (4.27)$$

concludes a result similar to (4.22), for the small region that the approximations are valid.

Comparing Figure 4-2 and Figure 4-5 as well Figure 4-3 and Figure 4-6, one can claim that in a three-microphone difference lens closeness function we have achieved a sharp and linear response compared to the two-microphone difference closeness function.

Notice that in Figure 4-2 as well Figure 4-5 the slope of the semi-linear and linear region varies with frequency. This in turn creates frequency dependent overall weightings.

In other words, by changing the frequency content of the signal and keeping the time average time frame constant, the slope of the e.g. linear region in Figure 4-6 changes. Since this overall slope change happens for all cells in the hemisphere in a similar way, the output of closeness function (though frequency dependent) maintains a viable

measure for closeness, taking into account the final linear averaging estimation strategy.

This property is valid for previous multiplicative based closeness function as well.

In this subsection, we defined a difference based closeness function (DCF) (4.13) and proved its linear response with angular deviation both in time and frequency domain.

The pseudo-frequency based DCF analysis helps us to study and visualize the behaviour of orthogonal microphone based closeness functions especially their linearity region versus frequency.

4.2 Correlative Closeness Function (CCF)

The idea of correlative based closeness function started from a simple fact that, in an endfire setting if we multiply a delayed version of the former microphone signal with the latter microphone signal, the outcome should be a peak high value while the sound source is exactly in line with the two-microphone direction. Heuristically one expects a high value equal to autocorrelation of the signal right in that direction and a decreasing RMS output versus sound source angular deviation.

4.2.1 Pinhole and Lens Correlative CF

Here we multiply both signals, instead of taking their difference. In this case, after time integration, we would have a correlation instead of difference error. The term:

$$C_{ef}^i(t) = S_0(t - T_r)S_i(t) \quad (4.28)$$

acts as a two-microphone pinhole detector. If the sound source is aligned to an end fire setting $S_0(t - T_r)$ and $S_i(t)$ will be highly correlated, i.e., $C_{ef}^i(t)$ has the maximum possible value in each time instance.

The endfire setting for m_0 and m_i occurs when m_0 and $m_{i_\perp}$ are in broadside setting.

Therefore if we calculate:

$$C_{bs}^i(t) = S_0(t - T_r) S_{i_\perp}(t) \quad (4.29)$$

The maximum for $C_{ef}^i(t)$ and $C_{bs}^i(t)$ occurs at the perpendicular source bearings (similar to other two CF cases).

Notice that we also can define another broadside term:

$$C_{bs-s}^i(t) = S_0(t) S_{i_\perp}(t) \quad (4.30)$$

Here the maximum occurs in the same direction as endfire term.

Considering (4.1), (4.2) and (4.28):

$$C_{ef}^i(t) = S(t - T_r) S(t - \tau_i) + n_0(t - T_r) S(t - \tau_i) + n_i(t) S(t - T_r) + n_0(t - T_r) n_i(t) \quad (4.31)$$

By combining all multiplicative noise terms in one term:

$$C_{ef}^i(t) = S(t - T_r) S(t - \tau_i) + N_{ef}(t) \quad (4.32)$$

$$N_{ef}(t) = n_0(t - T_r) S(t - \tau_i) + n_i(t) S(t - T_r) + n_0(t - T_r) n_i(t) \quad (4.33)$$

Note that the objective is to have the maximum output at the output of the closeness measure in addition to maximum noise cancellation. By choosing quadratic cost function, which resembles the noise power:

$$J(C_{ef}) = \|C_{ef}^i(t) - S(t - T_r) S(t - \tau_i)\|^2 = \|N_{ef}(t)\|^2 \quad (4.34)$$

The optimal maximum likelihood (ML) solution that minimizes J (Euclidian distance and noise power) is the sample mean or average [79]:

$$\hat{C}_{ef}^i = \arg \min[J(C_{ef})] = \overline{S(t - T_r) S(t - \tau_i)} \quad (4.35)$$

Similar reasoning can be applied to broadside term and therefore:

$$\hat{C}_{bs}^i = \arg \min[J(C_{bs})] = \overline{S(t-T_r)S(t-\tau_{i_\perp})} \quad (4.36)$$

To extract the three-microphone (lens) closeness function from endfire and broadside terms, similar to the previous methods we can divide the endfire term to broadside term:

$$\overline{C}_{CF}(t) = \frac{C_{ef}(t)}{C_{bs}(t)} = \frac{\overline{S_0(t-T_r)S_i(t)}}{\overline{S_0(t-T_r)S_{i_\perp}(t)}} \quad (4.37)$$

Using the optimally equivalent terms (4.35) and (4.36), for the numerator and denominator of (4.37):

$$\hat{C}_{CF}(t) \approx \frac{\overline{S(t-T_r)S(t-\tau_i)}}{\overline{S(t-T_r)S(t-\tau_{i_\perp})}} \quad (4.38)$$

Back to the Taylor approximation of $S(t-\tau)$ in (4.25), each of numerator or denominator can be written as:

$$C_{term} = \overline{S(t-T_r)S(t-\tau)} \approx \overline{S(t-T_r)[S(t-\tau_i)\dot{S}(t)]} = \overline{S(t-T_r)S(t) - \frac{\tau}{2}\{2S(t-T_r)\dot{S}(t)\}} \quad (4.39)$$

Similar to the previous analysis we change the statistical average with a practical limited time average. In the limit, we can replace this time average with the integral operation.

$$C_{term} \approx \overline{S(t-T_r)S(t)} - \frac{\tau}{2} \overline{\partial S^2(t-T_r)/\partial t} \quad (4.40)$$

$$C_{term} \approx R_{ss}(T_r) - \frac{\tau}{2} S^2(t-T_r) \quad (4.41)$$

In which $R_{ss}(T_r)$ is the autocorrelation of source signal at the delay point T_r .

Having (4.38) and (4.41), the correlative lens closeness function:

$$\hat{C}_{CF}(t) \approx \frac{1 - \frac{\tau_i}{2} \frac{S^2(t-T_r)}{R_{ss}(T_r)}}{1 - \frac{\tau_{i_\perp}}{2} \frac{S^2(t-T_r)}{R_{ss}(T_r)}} \quad (4.42)$$

However:

$$R_{ss}(T_r) \ll R_{ss}(0) = \overline{S^2(t)} = \overline{S^2(t-T_r)} \quad (4.43)$$

Therefore if:

$$\tau \gg \frac{2R_{ss}(T_r)}{S^2(t-T_r)} \quad (4.44)$$

we can neglect the one from numerator and denominator so:

$$\hat{C}_{CF}(t) \approx \frac{\tau_i \frac{S^2(t-T_r)}{2R_{ss}(T_r)}}{\tau_{i_\perp} \frac{S^2(t-T_r)}{2R_{ss}(T_r)}} \propto \frac{\tau_i}{\tau_{i_\perp}} = \frac{\cos(\theta)}{\sin(\theta)} = \cot(\theta) \approx \frac{1}{\theta} \quad (4.45)$$

This is the inverse linear relationship with deviation angle.

We claim that in most instances particularly when signal to noise ratio (SNR) is high, (4.44) is realizable.

In this subsection, we defined a correlation based closeness function (CCF) (4.13) and proved its linear response with angular deviation in the time domain. Also we demonstrated the necessary conditions to validate the linearity approximation.

4.3 Joined Closeness Functions

All closeness functions that we discussed so far have similar methodologies for both broadside and endfire terms.

During the course of our research we observed that one can enhance the result by joining different parts and styles of closeness function, at the cost of increasing the algorithm complexity. Two such approaches are discussed here:

1-Calculating two different closeness functions and taking the normalized average of both methods to enhance the closeness estimation. This method almost doubles the computational complexity of the whole algorithm.

2-Using altered methods for the broadside term and the endfire term. Since all endfire terms are cosine dependent and all broadside terms are sine dependent, one can use the broadside of one type of closeness function with the endfire of another. We explicitly tested the endfire of correlative closeness function with broadside of difference closeness function. Not to mention that for each case we may need to normalize each term to match the other:

$$JCF(t) = \frac{CCF_{ef}(t)}{DCF_{bs}(t)} = \frac{\overline{S_0(t-T_r)S_i(t)} / \overline{S_0^2(t)}}{\overline{S_0(t)-S_{i_\perp}(t)} / \overline{S_0(t)}} = \frac{\overline{S_0(t-T_r)S_i(t)}}{\overline{S_0^2(t)-S_0(t)S_{i_\perp}(t)}} \quad (4.46)$$

$$JCF(t) = \frac{R_{ss}(T) - \frac{\tau_i}{2} S^2(t-T_r)}{\frac{\tau_{i_\perp}}{2} S^2(t)} \approx \frac{\frac{\tau_i}{2} S^2(t-T_r)}{\frac{\tau_{i_\perp}}{2} S^2(t)} = \frac{\tau_i}{\tau_{i_\perp}} \approx \frac{1}{\theta} \quad (4.47)$$

considering similar approximations stated in subsection 4.2.1.

This way we could benefit from the higher dynamic range of the endfire correlative term and the better linearity of broadside difference term. Obviously, this is achieved at the price of increasing the computational cost.

4.4 Corollaries

Two new methods for closeness functions were discussed in this chapter in addition to one joint method. In this research we started our idea with difference closeness function (DCF) and correlative closeness function (CCF) first and later came up with the multiplicative closeness function (MCF) which was discussed thoroughly in the previous chapter. The MCF based eye array provided results with higher overall accuracy and lower outliers. The DCF linearity beat CCF but regarding its lower dynamic range, lack

of frequency spreading and its higher sensitivity to array's calibration the outcome is worse than other methods. The special version of joint closeness function that we discussed earlier, having CCF endfire and DCF broadside, has a good linearity and enhanced dynamic range and therefore better overall accuracy, but at the cost of increased computational complexity and lack of calculation similarity. Note that the computational simplicity and symmetric calculations were our main goals in this research.

5. Experimental Results

This chapter for the most part deals with the experimental results of our array system and algorithm, and its objective is to evaluate the system and compare the three associated categories of closeness functions in terms of coverage and accuracy in similar settings. The chapter starts with an illustration of the test environment and protocol for our experiments, followed by the results of our algorithm with MCF. Later, we discuss the result of the alternative closeness functions DCF and CCF which were introduced in Chapter 4. Subsequently the effect of the signal-to-noise ratio on the error is addressed. This chapter concludes with some remarks on the experimental results.

5.1 Test Protocols

In general the choice of measurement protocol depends on the application and analysis predictions. For each application, a good procedure has to be efficient, yielding measures of performance within a reasonable testing time. System accuracy is measured by the systematic difference between true and measured values. The accuracy measure, as will be described in Section 5.1.4, was chosen to be the RMS value of the angular error. The measurement precision, related to the error variability with repeated measurements was chosen to be the standard deviation of the angular error.

Preparing an absolute unified test environment to evaluate different algorithms and methods for the sound in enclosed areas is troublesome. This is due to the high variability of conditions in sound propagation such as different reverberation patterns, background noise, change of room temperature and humidity and therefore sound speed

and creates an uncontrolled test environment to deal with. To explain the results of our sound localization scheme, we tried to keep most of controlled test procedures intact. In these tests we set and track a measurement protocol to unify the circumstances for different variations in the algorithm.

5.1.1 Environment

The test acoustic environment is a $4.9 \times 3 \times 2.7$ m untreated office. Adverse acoustical condition in such a small room is noticeable [84]. Since we had no availability and access to anechoic chamber or acoustically treated rooms, we carried out the entire lab work and the final tests in the office which is an acoustically difficult environment. The room is not carpeted and it has a reverberation time of $RT60 \simeq 480ms$. In addition there are three computer desks and two bookshelves inside the room. The front and part of back of the room is covered with glass window and hard drapery. There is not much acoustically absorbing material in the room. The hemispherical microphone array was placed at one of the corner sides at the approximate height of 2.2 meters from the floor (Figure 5-1), on top of a movable tray, and the speaker was at the opposite corner of the room, attached to an adjustable stand, so that the source to array distance was approximately 2.5 meters. This distance is higher than the minimum distance required complying with the far field assumption [Appendix B].

Figure 5-1. The eye microphone array under test

5.1.2 Sound Source, Acquisition

The presence of large reflections increases the spatial variation within a sound field. The sound-pressure magnitude in response to a single tone signal in a reverberant environment would show large spatial derivatives corresponding to variations between peaks and troughs of the so called “standing waves”. This rapid spatial dependence complicates measurements of sound propagation within the room. In order to smooth out the spatial dependence and reduce the standing wave, one has to use either narrow-band noise or a chirp, in lab measurements [2].

Therefore, for all experiments, the test source signal was chosen to be a one second repeating chirp signal ranging from 100Hz to 4 kHz. The system was tested with other sound signals as well, e.g. band-limited noise and some repeating speech utterances.

To keep the sound source as omni-directional and wideband as possible, plus to reduce the chance of having a diffused sound field, we kept the loudness of the sound as low as a normal conversation 60 dBA SPL (Sound Pressure Level, referenced to 20 micropascals = 0 dB SPL [85]).

For a 4 kHz bandwidth, the minimum acceptable sampling frequency is 8 kHz according to the Nyquist theorem. However, the maximum possible sampling frequency of our data acquisition board (40 kHz per channel for 27 channels) was chosen in order to achieve the best angular resolution, as previously discussed in Subsection 3.6.1. Another way to increase the angular resolution is up-sampling of the Nyquist sampled signals. We attempted to upsample the 8 kHz sampled signals but the computational burden of upsampling filters of 27 channels was too high and prevented real time localization with our existing processing system (Subsection 7.3.4).

Since we had no apparatus to exactly localize the three dimensional spatial orientation of an adjustable sound source (speaker) relative to the spherical coordinates of the eye array, in addition to the spatial limitations of the test room, we rotated the array instead of the sound source in the majority of our tests. By employing a mechanical rotation system for azimuth and elevation along with installing two protractors with swinging arm indicators we made special goniometers (Figure 7-8) for the hemispherical array.

In the start of each test series, we fixed the array center direction to the sound source direction visually and reset the goniometers to zero degrees. To facilitate this procedure, we attached a laser pointer in front of the speaker. This way we can easily put m_0 and

m_1 in the direction of the laser beam. Subsequently each angular deviation of the array can be read by goniometers.

In most microphone array tests inside reverberant environments, one has to make sure that the array is in the area in which the direct signal is stronger than reverberation, for a reliable measurement. In general, acoustic measurements are inherently variable, e.g. small variations in microphone positions in the sound field or uncontrolled noise sources can cause significant variations in the measured sound signal. Furthermore the tester's presence and the presence of others in the room where you make measurements, as well as outside noise, will increase the uncertainty associated with any sound measurement. Therefore a well thought out plan of repetitions of measurements is necessary to determine the variability of the estimates and it allows us to distinguish among uncontrolled room noise and parameter variations with constant errors such as microphone array biases.

5.1.3 Visualization

In order to visually illustrate the closeness function values for each direction and compare them with each other, we constantly monitored their value in real-time plots. Therefore we visualize the closeness function value of each direction with a greyscale point on the microphone locations projected on a plane. The high closeness function values resemble brighter points. In these illustrations, since we have projected the microphone locations on the hemisphere onto a plane, due to fisheye effect, the barrel distortion squeezes the corner microphone locations. Instances of these closeness function pictures are good visual representations to discover the closeness function

behaviour. Figure 5-2 (left) shows a sample of this closeness function illustration with all CF points brighten along with microphone numberings at the right side.

Figure 5-2 Closeness function visualization with microphone numbering

5.1.4 Error Calculation

In the following, error of the sound source direction estimation is defined as the angular difference between the actual direction and the estimated direction.

Consider the i^{th} targeted angle of the sound source to be $[\varphi_s^i, \theta_s^i]$ and its corresponding estimated angle as $[\varphi_e^i, \theta_e^i]$. Therefore, the spherical angular distance between these two directions is the angle subjected to the distance of the shortest path along the surface of the unit sphere and always lies along a great circle. The corresponding unit vectors for source angle and estimated angle in Cartesian coordinate are as follows:

$$\mathbf{v}_s^i = [\sin(\varphi_s^i) \cos(\theta_s^i) \quad \sin(\varphi_s^i) \sin(\theta_s^i) \quad \cos(\varphi_s^i)]^T \quad (5.1)$$

$$\mathbf{v}_e^i = [\sin(\varphi_e^i) \cos(\theta_e^i) \quad \sin(\varphi_e^i) \sin(\theta_e^i) \quad \cos(\varphi_e^i)]^T \quad (5.2)$$

The shortest angular distance between $\mathbf{v}_s^i$ and $\mathbf{v}_e^i$ is:

$$\varepsilon^i = \arccos(\nu_s^i \cdot \nu_e^i) \quad (5.3)$$

But from (5.1) and (5.2):

$$\begin{aligned} \nu_s^i \cdot \nu_e^i &= \cos(\varphi_s^i) \cos(\varphi_e^i) + \sin(\varphi_s^i) \sin(\theta_s^i) \sin(\varphi_e^i) \sin(\theta_e^i) + \sin(\varphi_s^i) \cos(\theta_s^i) \sin(\varphi_e^i) \cos(\theta_e^i) \\ &= \cos(\varphi_s^i) \cos(\varphi_e^i) + \sin(\varphi_s^i) \sin(\varphi_e^i) [\sin(\theta_s^i) \sin(\theta_e^i) + \cos(\theta_s^i) \cos(\theta_e^i)] \\ &= \cos(\varphi_s^i) \cos(\varphi_e^i) + \sin(\varphi_s^i) \sin(\varphi_e^i) \cos(\theta_s^i - \theta_e^i) \end{aligned} \quad (5.4)$$

Therefore:

$$\varepsilon^i = \arccos[\cos(\varphi_s^i) \cos(\varphi_e^i) + \sin(\varphi_s^i) \sin(\varphi_e^i) \cos(\theta_s^i - \theta_e^i)] \quad (5.5)$$

Here ε^i is the angular distance (error) between i^{th} targeted sound source and its estimation. This angular distance is positive $\varepsilon^i \geq 0$, therefore both average and RMS can be considered as measures of error. To comply with some of the research in this area [1, 2], we decided to calculate the RMS of error for a solid angle by:

$$\bar{\varepsilon} = \sqrt{\frac{1}{N} \sum_{i=1}^N \varepsilon^i{}^2} \quad (5.6)$$

5.2 Results Utilizing MCF

Each time, we started the test process with the sound source directed to $\theta = \varphi = 0$ degrees and rotated the array in equal segments of 5 degrees in both azimuth and elevation directions to cover the whole ± 90 degrees span. At directions close to the corners of the hemisphere, the microphone omni-directionality is highly uncertain due to the encapsulation in the array joint microphone holders. We carried out the tests even for the borders which either azimuth or elevation is close to ± 90 degrees. Due to some practical mechanical limitations in rotation of the array in elevations of $45^\circ \leq \varphi \leq 90^\circ$ (array stand blocking the array rotation mechanism), we rotated the array with its stand in intervals in that area and measured the elevation angle with an extra protractor

installed on the tray. We intentionally ignored some of the points, mostly in $45^\circ \leq \varphi \leq 90^\circ$ range, due to inaccuracy of measurement of the elevation angle. Later, we filled the neglected points by interpolating with adjacent measurements. The total number of the directions that we tested was 1226. Afterwards we computed the estimation error by calculating the angular distance between the measured direction and the estimated direction based on (5.5) and (5.6) in Subsection 5.1.4.

The estimation algorithm is similar to Subsection 3.9 and Appendix A, with a set integration time of 1s. The integration time (averaging time) is practically equal to the sampling time T_s multiplied by the number of samples in a frame, N in (A.1). Obviously, a longer averaging time frame would result in enhanced localization accuracy at the price of boosting the response lag. We chose a high integration time (1s.) to guarantee the best result with the existing settings.

Figure 5-3 illustrates one sample snapshot of closeness function values for a specified direction. Here the sound source targets the point $(\varphi_s = -45^\circ, \theta_s = -15^\circ)$ which is somewhere inside the triangle $\triangle m_5, m_6, m_{14}$.

As illustrated, the closeness functions related to microphones m_5, m_6 and m_{14} have the maximum values (brighter points) and some other closeness functions which are neighboring that area have lower values.

Figure 5-3 Closeness function snapshot (X is the sound source direction)

The closeness functions which are far from the source area have absolutely lower values which are shown with dark spots.

As discussed in Subsection 3.9.1 and Appendix A, k of the highest valued closeness functions must be selected before the final estimation, moreover, theoretically the minimum value for k is 3. Choosing $k = 3$, we occasionally observed that moving the sound source such that its direction vector crosses the border from one spherical triangle to a neighboring spherical triangle, a stepwise change is experienced in the final estimated direction. This sharp jump in estimation is mainly due to the imbalance in microphone signal gains. Theoretically, there should be no abrupt alteration due to toggling one of the estimation directions with another, but in practice we experienced some abrupt changes due to the constant gain imbalance and the lack of omnidirectionality.

Thus we decided to increase the number of final estimator cells. Choosing $k = 4$, the abrupt changes were lessened but they still were present. With the final choice of $k = 5$, the abrupt changes were not present anymore, at least in the majority of central area of the swath.

In the example related to Figure 5-3, to calculate a step free estimate, the algorithm also utilizes closeness functions related to m_{13} and m_{15} .

To acquire a better sense, the following figures illustrate different snapshots of the closeness function values, with the exact sound source direction indicated on them with a cross sign.

Figure 5-4 Four different snapshot of MCF with source directions

As you can observe from Figure 5-4, the higher valued closeness functions are always in the vicinity of the sound source direction when we use MCF. If we choose a shorter averaging time, e.g. 50-100 ms, occasionally some of the high valued closeness functions fall far away from the sound source direction. These outlier closeness functions degrade the final estimation, depending on their rank and value. For MCF, the proper selection of the averaging time prevents the outlier issue to a high extent.

Figure 5-5 is a three dimensional illustration of calculated error of the system utilizing MCF versus azimuth and elevation angles. The angular spacing between every measurement is five degrees.

Figure 5-5 The MCF error versus azimuth and elevation

As illustrated, the error rate increases and varies rapidly when the sound source approaches the border of the array. In addition, the error is predominantly lower and

steadier in the central area of the swath angle. This can be seen by observing the error variations through a quiver plot. A quiver plot (also known as velocity plot) is a plot type that shows a three dimensional graph in a plane by showing the third dimension (here, error gradient) as vectors originating from measurement points. It is mostly useful in the 3D graphs with patterned gradient values. The quiver plot of gradient error vectors $q^i = [\partial \varepsilon^i / \partial \theta_s^i, \partial \varepsilon^i / \partial \phi_s^i]^T$ for MCF is shown in Figure5-6. As shown the error changes are increasing at the corners.

Figure 5-6 Quiver plot of MCF gradient error versus azimuth and elevation

Theoretically the eye array and algorithm ought to cover the whole ± 90 degrees of azimuth and elevation, but practically at higher angles of azimuth and elevation in which

the sound source falls into triangles in the outer layer of the array, the error increases significantly.

The primary reason for this rapid increase of error is the choice of $k = 5$ in the final estimation, which forces our algorithm to select two additional cells (compared with $k = 3$) which are towards the inside of the array, in situations where the source targets the border layer of triangles. This pushes the estimated source direction toward the center of the array due to unavailability of shell directions beyond the final layer borders ($\varphi = \pm 90^\circ$ or $\theta = \pm 90^\circ$). We anticipate this error could be reduced by adding another extra layer of triangles to the hemisphere and establishing an extended hemispherical array. As well, choosing $k = 3$ while the sound source targets the border layer of triangles can reduce the effect of this error, at the cost of adding extra logical expressions at the last stage of the algorithm, which is contrary to our initial goals. As stated in our initial statement of purpose, we seek solutions which are simple and realizable with an analog or mixed signal integrated circuit. Therefore, we intentionally rejected the use of complex logical filtering or neighbor based clustering in our algorithm.

Contrary to microphones in the central direction, in border direction microphones, the levels of omni-directionality of most of the microphones are questionable due to encapsulation inside their joint holders (Figure 7-6). This is the secondary reason for the ascending shape of the error in the corners.

Figure 5-5 and Figure 5-6 show that a semi-circular area in the center has lower error compared to the rest. If we magnify the center arrows of Figure 5-6 and remove arrows of $65 \leq |\varphi| \leq 90$ or $65 \leq |\theta| \leq 90$, we can see that (Figure 5-7) beyond a circle with radius of about 65 degrees, the error arrows begin to enlarge, generally directed toward outside the circle.

Therefore the working region of our system with MCF can be considered as a circular area with radius 65 degrees in the azimuth vs. elevation plane.

For the present system (without adding an extra layer of microphones and with $k = 5$), the system practically tracks any sound source within that circle, with reasonable accuracy.

The RMS error of the circular area is 3.07 degrees (with a standard deviation of 0.69 degrees). The RMS error of the whole $\pm 90^\circ$ swath is 7.8 degrees, however having errors up to 27 degrees in outer triangle layer, disqualifies those directions as acceptable swath range.

Figure 5-7 Quiver plot of central area of the MCF gradient

The accuracy of any array based algorithm is dependent on its aggregated sensor placement. Since our test structure was fabricated using polyethylene plastic rods, the whole structure easily warped slightly and lost its initial hemispherical shape to some

extent after some time. Therefore the current distance between the central microphone and the shell varies from 33 to 36cm. This is another source of estimation error.

The RMS error plot for the horizontal direction (the line with zero elevation) and vertical direction (the line with zero azimuth) are drawn in Figure 5-8.

As illustrated, the range of low error is slightly different in both directions. In other words, the circular shape of low error slightly tends to oval. We speculate three probable reasons for this:

- The slightly distorted structure of the array.
- Different measurement accuracies for azimuth and elevation, due to mechanical limitation since we measured part of the elevation angle with an external protractor which might add inaccuracy to our elevation measurement and mechanical error propagation in that direction.
- Slightly different patterns of reverberated sound comparing higher azimuth angles with higher elevation angles.

Figure 5-8 MCF vertical and horizontal error plot

5.3 Results utilizing DCF

Figure 5-9 is a three dimensional illustration of calculated error of the system utilizing the difference closeness function versus azimuth and elevation angles. The angle spacing of measurement is five degrees in both azimuth and elevation directions.

Figure 5-9 The DCF error versus azimuth and elevation

Also the quiver plot of DCF gradient error is shown in Figure 5-10.

Similar to MCF, in DCF the error has a relatively flat shape in central areas and soaring profile in outer layer. Nevertheless as expected in both areas the level of error is much higher than with the MCF.

Figure 5-10 Quiver plot of DCF gradient error versus azimuth and elevation

Looking at the closeness function pattern, one can observe that, unlike the MCF, the DCF experiences more outliers among closeness functions, particularly at the corners. Figure 5-11 shows a snapshot of the DCF closeness function which contains outliers mostly on the opposite side of the sound source direction. Besides, the number of unsystematic outlier occurrences in the middle area is higher than the MCF case.

Note that the occurrence of an outlier does not necessarily mean that the outlier negatively affects the output result. To have a negative effect, the outlier value has to be sufficiently high to fall into the top k closeness functions that estimate the sound source direction.

Figure 5-11 A sample snapshot of DCF values with visible outliers

Similar to the MCF case, by neglecting the result for $65 \leq |\varphi| \leq 90$ or $65 \leq |\theta| \leq 90$ and magnifying the central arrows, we can observe that, yet again the low error area has a circular shape. But here the arrows start to enlarge from nearly 60 degrees.

Figure 5-12 Quiver plot of central area of DCF gradient

However we took the RMS error for the same circle as in MCF case with a radius of 65 degrees in order to have the same comparison region. Here the RMS error for a 65 degree circle is 6.09 degrees which is nearly twice that of the MCF case. Moreover, the standard deviation of error in the circular area is 1.77 degrees which shows that the error variation is 2.5 times the error variation in MCF case. Further, the RMS error of the whole $\pm 90^\circ$ swath is 14.55 degrees and the maximum error value in outer triangle layer is 57 degrees.

Similarly, the error plot for the horizontal direction (the line with zero elevation) and vertical direction (the line with zero azimuth) are drawn in Figure 5-13.

Figure 5-13 DCF vertical and horizontal error plot

As stated earlier, the reason for high error of DCF compared to MCF is mostly the existence of outliers. A reason for production of outliers is the adverse effect of additive noise, reverberation and channel misalignments. Calibration of microphone arrays is troublesome and any mismatch influences the outcome of any subtractive term while the multiplicative terms are more robust to small perturbations in amplitude[81]. Therefore mismatching creates closeness functions which are biased and have lower dynamic ranges in DCF in comparison to MCF. Low dynamic range CFs are more prone to false selection in the final stage of the algorithm.

In other words, DCF with additive noise is more susceptible to false information than MCF with multiplicative noise (Subsection 3.6.3).

Another reason for production of outliers in DCF is theoretical and due to lack of spectral spreading in DCF compared with MCF. The spectral spreading of multiplicative signals and noise [81, 86] can enlarge the frequency range and fill up the frequency gaps in nonlinear regions that have closeness functions comparable to linear region closeness function values hence outliers decrease in MCF.

5.4 Results Utilizing CCF

Figure 5-14 is a three dimensional illustration of calculated error of the system utilizing Correlative Closeness Function (CCF) versus azimuth and elevation angles. The angle spacing is five degrees in both azimuth and elevation directions.

Figure 5-14 The CCF error versus azimuth and elevation

Also the quiver plot of CCF gradient error is shown in Figure 5-15.

Similar to MCF and DCF, the CCF error also has a relatively flat shape in central areas and soaring error in outer layer. The error level is not as high as DCF and not as low as MCF.

Figure 5-15 Quiver plot of CCF gradient error versus azimuth and elevation

Generally the CCF behaviour in CF snapshots is close to MCF. The corner outliers that were experienced in DCF are uncommon here in CCF.

By neglecting the result for $65 \leq |\phi| \leq 90$ or $65 \leq |\theta| \leq 90$ and magnifying the central arrows, we can observe that, yet again the low error area has a circular shape (Figure 5-16).

If we took the RMS error for the same circle as in MCF and DCF cases with a radius of 65 degrees, the RMS error is 4.15 degrees which is nearly 70% of DCF case and 135% of the MCF case. The RMS error of the whole $\pm 90^\circ$ swath is 9.5 degrees and the maximum error value in outer triangle layer is 55 degrees. The standard deviation of error is 1.35 degrees, which is 76% of DCF and 195% of MCF case.

Figure 5-16 Quiver plot of central area of CCF gradient

Similarly, the error plot for the horizontal direction (the line with zero elevation) and vertical direction (the line with zero azimuth) are drawn in Figure 5-17.

Figure 5-17 CCF vertical and horizontal error plot

The CCF was the next closeness function that we worked on after DCF in this research. The advent of multiplication, softened the result as well reduced the error rate. The reason that MCF outperforms CCF is likely due to two factors:

- The MCF formula in its open form contains more multiplications than CCF. Therefore we hypothesize that it might cover broader frequency spreading compared with CCF, and therefore better overall result.
- Furthermore, in CCF with the assistance of a Taylor approximation (4.39), an integration supposition (4.41) and an extra assumption (4.44) we showed its cotangent behaviour hence its inverse linear relationship with deviation angle. While in MCF, the correspondence of MCF and cotangent of deviation angle is

shown without any simplification. In other words, the likeliness of CCF with cotangent presumes some extra pre-conditions which might not be met always. Therefore MCF is more prone to behave as an essentially inverse linear relationship as compared to CCF.

5.5 Effect of the Signal-to-Noise Ratio

We consider the noise variance as the average of microphone signal variance in the silent room (no speaker signal). The signal variance is the average of microphone signal variance while the speaker is on, minus the noise variance (speaker off). Therefore we can have a practical measurement based definition for signal-to-noise ratio (SNR). For each segment of the experimental data collected, the average SNR across all microphones in the array was computed using the following formula:

$$SNR = 10 \log_{10} \left(\frac{\sum_{i=1}^{27} \sigma_{s_i}^2}{\sum_{i=1}^{27} \sigma_{n_i}^2} - 1 \right) \quad (5.7)$$

where $\sigma_{s_i}^2$ is the variance of the signal acquired from m_i with the speaker on, and $\sigma_{n_i}^2$ is the variance of the signal acquired from m_i with the speaker off. We also measured the zero signal sound level and finite signal sound level with an Integrating Sound Level Meter (ISLM) made by Bruel & Kjaer®, type 2225 (Figure 7-17), installed close to the array, and compared the SNR acquired with it with our microphone array measurement. To observe the effect of increasing the noise on the measurement error, we setup another set of experiments. We placed a subwoofer and a speaker under a computer desk to create a diffuse (or semi-diffuse) noise field via signal side emission subsequent to reverberation on parallel planes [87]. Applying a low pass filtered white noise to this

system, we controlled the noise level of the room and therefore reduced SNR and performed some acquisition with 9 different SNR levels in 13 different directions (10 degree spacing in azimuth plane). The average noise level inside the room close to the array measured by ISLM while the computer is off is 32 dBA. When the computer is turned on, the room noise level varies from 42 dBA to 47 dBA (depending on the extra sound due to computer hard drive activity). Considering 44 bBA average noise level in array location and comparing it with the adjusted 60 dBA test sound level (speaker on), the average SNR of all our previous tests is 16 dB, which is the minimum SNR in typical situation of a conference room. We first increased the noise in 6 steps of 3 dB. Later by returning to the normal noise level, we increased the signal level in two steps of 3 dB to experience higher SNR. The overall results are depicted in the graph showing RMS error versus SNR for all closeness functions in Figure 5-18.

Figure 5-18 Average error performance versus SNR

As you can see from the graph, by decreasing the SNR the average error increases. This increase is minor in $\text{SNR} > 4$ dB and after that the performance degrades rapidly. We also notice that the signal variance and therefore SNR acquired by the microphone array distinctly reduces on higher sound source angles. This most likely is due to deviation of the microphones from omni-directionality because of their encapsulation inside the microphone joint holders. The lack of omni-directionality affects the overall signal variance when the sound source is targeting borders. Conversely when the sound source targets inside the array, most of microphones face the sound and the encapsulation has less effect. In the graph of Figure 5-18 we considered the SNR acquired by microphone array, which is affected by lack of omni-directionality. But since we limited the error averaging on the results amid $-65^\circ \leq \varphi \leq 65^\circ$ (the working area of our array), the deviated SNR does not have much influence on the final results.

Unexpectedly, the DCF performed better than CCF in low SNR values. With all three closeness functions, the average error increased slightly in $\text{SNR} = 22$ dB. We speculate this is due to the increase of sound diffuseness due to high signal levels in enclosures with hard surfaces [87].

Overall with the current setting and implementation, our system performs well, above a SNR of 4 dB (Figure 5-18).

5.6 Array Size, System Noise, Outliers, Number of Cells

If a pair of microphones are getting close to each other, there is more possibility for them to experience similar background, propagation channel, orientation from source and reverberation patterns. This implies a higher likelihood of dependence or correlation between such signals. Therefore they may experience less outlier (anomalies) in their

pairwise cross correlation or closeness function patterns. In other words, spatial comparative processing based on closely spaced pairs generally achieves higher “correct” outputs. However, closely spaced microphones experience very small time delays even if sound source is in the endfire position. Thus the sampling and quantization error has more adverse effect on the output. On the contrary, if a pair of microphones is widely separated from each other, their signals have higher probability to get different reverberations, so their final spatial process experience higher outliers or lower “correct” output, although their time difference is longer and their sampling and quantization related error is lower.

Using a huge microphone array with 448 microphones in a $8m \times 8m \times 3m$ room, the research in [26] experimentally shows that for microphone pair distances less than 40cm, the average likelihood of correct cross correlations increases sharply by reducing the distance, for distanced between 40cm-100cm the average likelihood of correct cross correlation stays fairly constant in all SNRs, and for distances more than 100cm it sharply reduces with increasing the distance. They also show the increase of quantization error with reducing the microphone distance.

Although shrinking the array size reduces the adverse effect of reverberation, it may increase the effect of the noise caused by sampling and quantization. Therefore any reduction in size has to be followed by enhancing the digital system or utilizing an analog implementation.

Shrinking the array size also broadens the angular range of linearity for all closeness functions. This lets us use larger number of closeness functions in the final estimation algorithm (increasing k). Lowering the maximum frequency content of the incoming

signal, though bounds the functionality of the system to low frequencies sources, it also broadens the closeness function linear range (Refer to Figure 4-2 and Figure 4-5).

One can benefit from either or both of these techniques to increase the linear region of closeness functions and therefore increase the number of cells involved in the final estimation. As discussed both methods have their own known drawbacks.

5.7 Remarks

We checked our algorithm and array with three different types of closeness functions and assessed them in experiments with similar settings. The result explains that, although having different errors average values, almost all three types of closeness functions have similar error pattern and coverage. The results are summarized in Table 5-1.

Table 5-1 Error and coverage versus closeness function

Closeness Function	Coverage (steradians)	RMS Error (degrees)
DCF	$\frac{4}{3} \pi$	6.09
CCF	$\frac{4}{3} \pi$	4.15
MCF	$\frac{4}{3} \pi$	3.06

However DCF and CCF have less computation load than MCF.

In another set of experiments we observed a mild and similar error increase with decreasing SNR>4dB for all CFs plus an enhanced performance for DCF in very low SNRs.

6. Refinement of the Approach:

Eye Array Placement

This chapter starts with the problem of placement of the eye array in an enclosed area, continues with lens cell sensitivity and trihedral corner characteristics, and concludes with a suggested placement of an eye array which reduces the adverse effect of the reverberation. Anechoic environments and damped walls reduce the power of reverberated sound, while in this placement strategy we diminish the negative effect of the reverberation on our array.

6.1 Where to Place the Array?

The placement of microphone arrays is very important. The major problem in an enclosed area is reverberation. The importance of reverberations in an acoustic environment is primarily dependent on the sound absorptive properties of the boundaries and furnishings of the environment. Anechoic environments are bounded by surfaces that are designed to be matched to the characteristic impedance of air. Those surfaces absorb all incident sound waves, such that there are no reflections. Rigid boundaries absorb little sound energy and cause large reflections.

Perhaps one of the oldest documented scientific perceptual designs towards utilizing reflective properties of sound in architectures is Kircher's drawings (Figure 6-1) in Acoustics section of his book in 1650 [88].

Figure 6-1 Kircher's acoustical perception; the emergence of reflection and echoes [40]

The reverberation in an enclosed area can be divided into two main parts:

- Early reverberation: Part of the reverberation that reaches the listener/microphone relatively early. The strengths of the reflections in early reverberation are high. In addition, the number of the paths is limited; therefore the early reverberation is quite directional.
- Late reverberation: The part of the reverberation that appears after early reverberation. Because of numerous reflections and absorptions, the strengths are low as well the directionality tends toward pseudo-omni-directional (uniform) during the decay process, especially in the middle of the enclosed area [89].

The late reverberation level is usually higher in the interior of the enclosed areas compared to the corners. There is a rule of thumb among sound specialists to record

beside corners in order to escape the adverse effect of reverberation, especially in the large enclosed areas [90].

Another key factor in placement of the eye array is the ability to encompass the entire three dimensional space.

Since our array has a hemispherical shape, the space inadequacy can be another factor. The upper trihedral corners of an enclosed area can be the best choice when it comes to space restriction and total area exposure.

Further in this chapter, we will show that based on the lens cell directional sensitivity, symmetry of the array and the retroreflection property of a trihedral corner, the upper trihedral corner of an enclosed area can diminish the negative consequences of early reverberation.

6.2 Lens Cell Sensitivity to Reverberation

Here we would like to analyze the sensitivity of our lens cell to the foremost reflection direction. Let us consider that we have a planar reflected sound with a deviation angle of θ_r from the main direction of the lens cell (Figure 6-2).

Figure 6-2 Lens cell with foremost reflection

The signal added with one foremost reflection is:

$$\tilde{S}_0(t) = S_0(t) + S_{r0}(t) \quad (6.1)$$

$$\tilde{S}_i(t) = S_i(t) + S_{ri}(t) \quad (6.2)$$

$$\tilde{S}_i^\perp(t) = S_i^\perp(t) + S_{ri}^\perp(t) \quad (6.3)$$

Without loss of generality, we can transfer the time origin of the reflected part to the center microphone. Therefore:

$$S_{r0}(t) = S_r(t) \quad (6.4)$$

$$S_{ri}(t) = S_r(t - \tau_{ri}) = S_r(t - \cos(\theta_r)T_r) \quad (6.5)$$

$$S_{ri}^\perp(t) = S_r(t - \tau_{ri}^\perp) = S_r(t - \sin(\theta_r)T_r) \quad (6.6)$$

Therefore the lens cell multiplicative closeness function (3.22) can be modified as:

$$CF_i = \frac{\tau_i}{\tau_i^\perp} = \frac{\langle [\tilde{S}_0(t) - \tilde{S}_i(t)] \cdot [\tilde{S}_0(t) - \tilde{S}_0(t - T_r)] \rangle}{\langle [\tilde{S}_0(t) - \tilde{S}_i^\perp(t)] \cdot [\tilde{S}_0(t) - \tilde{S}_0(t - T_r)] \rangle} \quad (6.7)$$

Assuming the sound source and room situation (reverberated path) is constant during integration time, we can temporarily ignore the averaging sign:

$$CF_i|_t = \frac{[\tilde{S}_0(t) - \tilde{S}_i(t)] \cdot [\tilde{S}_0(t) - \tilde{S}_0(t - T_r)]}{[\tilde{S}_0(t) - \tilde{S}_i^\perp(t)] \cdot [\tilde{S}_0(t) - \tilde{S}_0(t - T_r)]} \quad (6.8)$$

By placing equations (6.1) to (6.6), simplifying and collecting the θ_r related parts together, we can model the result likewise:

$$CF_i|_t = \frac{N1 + N2 \cdot [S_r(t) - S_r(t - \cos(\theta_r)T_r)]}{D1 + D2 \cdot [S_r(t) - S_r(t - \sin(\theta_r)T_r)]} \quad (6.9)$$

Consider two extreme cases of $\theta_r \approx 0$ (back reflection) and $\theta_r \approx \pm 90^\circ$ (side reflection):

$$CF_i|_t = \frac{N1 + N2[S_r(t) - S_r(t - T_r)]}{D1} \quad \forall \quad \theta_r \approx 0 \quad (6.10)$$

$$CF_i = \frac{N1}{D1 + D2[S_r(t) - S_r(t - T_r)]} \quad \forall \quad \theta_r \approx \pm 90^\circ \quad (6.11)$$

The reflected signal S_r is considered to be weaker than the main signal and therefore the term $[S_r(t) - S_r(t - T_r)]$ is small.

If we consider the case that the main signal is in close vicinity of the cell direction, the numerator of the closeness function is much higher in value than the denominator:

$$N1 \gg D1 \quad \forall \quad \theta_i \text{ small} \quad (6.12)$$

In this case, the equation (6.10) is not as much influenced by θ_r than equation (6.11). In other words, while our closeness function numerator term is high and denominator is low, the overall will not be affected much by a disturbance on the nominator than a disturbance on the denominator.

Therefore if the sound source direction is close to the lens cell direction (θ_i small), the cell output is not greatly affected by back ($\theta_r \approx 0$) as by side reverberations ($\theta_r \approx \pm 90^\circ$).

6.3 Trihedral Corner Retroreflection Property

Consider a trihedral corner with three orthogonal specular reflective surfaces (Figure 6-3). Given the far field assumption, we consider a planar sound wave S (normalized plane vector $\hat{s}$) hits one of the surfaces.

$$\hat{s} = \sin(\theta) \cos(\varphi) \hat{x} + \sin(\theta) \sin(\varphi) \hat{y} + \cos(\theta) \hat{z} \quad (6.13)$$

Suppose the signal first hits the z-y (L1) plane. Later, the reflection from z-y (L1) hits x-y (L2) plane and finally the reflection from z-y (L1) and x-y (L2) hits x-z (L3) plane.

Without loss of generality we can reorder the hitting sequence given the symmetry in trihedral shape.

Figure 6-3 Orthogonal trihedral corner

Normalized vectors of L1, L2 and L3 planes are consequently $\hat{x}$, $\hat{y}$ and $\hat{z}$.

Given $\hat{p}$ as the normalized vector of any specular surface, the vector form of geometrical law of reflection is as follows:

$$\cos(\vartheta) = \hat{s} \cdot \hat{p} \quad (6.14)$$

$$\hat{s}_{\text{reflected}} = \hat{s}_{\text{incident}} - 2 \cos(\vartheta) \hat{p} \quad (6.15)$$

Here, ϑ is the smaller of the angles between the incident wave $\hat{s}$ and the plane norm $\hat{p}$.

Therefore the reflected sound direction from plane L1 is:

$$\hat{s}_{L1} = \hat{s} - 2(\hat{s} \cdot \hat{x})\hat{x} = \hat{s} - 2 \sin(\theta) \cos(\varphi) \hat{x} \quad (6.16)$$

$$\hat{s}_{L1} = -\sin(\theta) \cos(\varphi) \hat{x} + \sin(\theta) \sin(\varphi) \hat{y} + \cos(\varphi) \hat{z} \quad (6.17)$$

As well, the reflected sound direction from L1 then L2 is:

$$\hat{s}_{L1L2} = \hat{s}_{L1} - 2(\hat{s}_{L1} \cdot \hat{y})\hat{y} = \hat{s}_{L1} - 2 \sin(\theta) \sin(\varphi) \hat{y} \quad (6.18)$$

$$\hat{s}_{L1L2} = -\sin(\theta) \cos(\varphi) \hat{x} - \sin(\theta) \sin(\varphi) \hat{y} + \cos(\theta) \hat{z} \quad (6.19)$$

Similarly, the overall reflection direction from all three planes L1, L2 and L3 is:

$$\hat{s}_{L1L2L3} = \hat{s}_{L1L2} - 2(\hat{s}_{L1L2} \cdot \hat{z})\hat{z} = \hat{s}_{L1L2} - 2\cos(\theta)\hat{z} \quad (6.20)$$

$$\hat{s}_{L1L2L3} = -\sin(\theta)\cos(\varphi)\hat{x} - \sin(\theta)\sin(\varphi)\hat{y} - \cos(\theta)\hat{z} = -\hat{s} \quad (6.21)$$

Therefore the overall reflection from a trihedral corner is in the reverse direction of the incident wave. This property is called retroreflection. If one throws a small ball into the corner of a room, it would return back the same direction after bouncing off the three surfaces of the orthogonal corner. Retroreflection of trihedral corners has been known for years in optics and radar hence there are lots of applications in those areas. The inside corner of a mirror-coated corner reflector sends the light back parallel to its original path. If one points a thin beam of laser light right near the corner, the beam would bounce from mirror to mirror and then exit parallel to the entering beam. Symmetrical arrays of corner mirror reflectors are used to make safety reflectors for cars, bicycles, and signs.

6.4 Upper Trihedral Corner Placement

There are some reports regarding ultrasonic applications of dihedral corners for Robot tracking [91]. Audio wavelengths are usually believed to be longer than the wavelengths that allow the use of practical size array corner reflectors. However, when it comes to the typically large and already constructed corner of an enclosed area such as a room, we only need to maintain enough distance to satisfy the far field assumption.

There are four orthogonal trihedral corners in the upper part of the majority of the enclosed areas. Those corners are generally unfilled. In view of the facts that eye array sensitivity to back reflections is minimum as well orthogonal corners have the retroreflection property; these corners can be the privileged candidates for eye array placement (Figure 6-4). An extra bonus for such a placement is the upper corner's

maximum visibility (coverage) of the entire three dimensional area. Also, if one chooses to cover the rear region of the eye array (corner walls) with acoustic damping materials intended for extra reduction of reflections, it will not be much evident and therefore will not influence the room decoration.

Figure 6-4 Eye array placement in upper trihedral corner

6.5 Evaluation

To evaluate our new placement strategy, we compared the results of our eye array in two distinct locations in a non-treated small office room. The first location was a dihedral corner and the latter was the middle of a side wall. Due to the difficulty of proper installation of our eye array to an upper trihedral corner, we decided to test the idea with a dihedral corner. Therefore we tested our system in a plane perpendicular to dihedral corner separation line, in order to maintain the retroreflection property. Note that a dihedral corner has the retroreflection property if the wave direction falls inside a horizontal plane. As you may ponder, in dihedral corner installation, although the

elevation swath range is still -90° to $+90^\circ$, the azimuth swath reduces to -45° to $+45^\circ$, due to physical limitation by dihedral corner walls. Thus we moved the sound source azimuth direction from -40° to $+40^\circ$ in 5 degree steps, and measured the root mean squared (RMS) error for both placement configurations. The result is depicted in (Figure6-5).

Figure 6-5 RMS error versus source bearing angle for dihedral corner and single wall placements

As illustrated the retroreflection property of the dihedral corner has created a smooth and small error compared with the single wall case.

For trihedral corner placement, the eye array has to be mounted so that its central axis is on the trihedral corner's symmetry axis (boresight axis). One can easily prove that this line is on $\theta = 54.7^\circ$ and $\varphi = 45^\circ$. In this case, the corner limits the elevation swath range to -54.7° to $+35.3^\circ$ and the azimuth swath to -45° to $+45^\circ$. This is a coverage which maps the central area of our array and therefore allows the system to work on higher accuracy regions.

6.6 Multiple Array Localization and Placement

In certain environments, however, multiple microphone arrays may be operating. Integrating the results of these arrays might result in a more precise sound localization system than that obtained by a single array. Furthermore, in large environments, multiple arrays are required to cover the entire space of interest. In these situations, there will be regions in which multiple arrays overlap in the localization of the sound sources [92]. Here, we are concerned with estimating the location of a wideband source using multiple sensor arrays that are distributed over an area. We considered schemes that distribute the processing between individual arrays and a fusion centre in order to limit the communication bandwidth between arrays and the fusion centre.

Triangulation is a standard approach for source localization with multiple sensor arrays. Each array estimates a bearing and transmits the bearing to the fusion centre, which combines the bearings to estimate the source location. Triangulation is characterized by low communication bandwidth and low complexity, but it ignores coherence that may be present in the wave fronts that are received at distributed arrays. There are two argued methods [93]:

- Ordinary triangulation, where each array estimates the source bearing and transmits the bearing estimate to the fusion centre. This approach does not exploit wave front coherence between the distributed arrays, but it minimizes the communication bandwidth between the array and the fusion centre.
- Each array estimates the source bearing and transmits the bearing estimate to the fusion centre. In addition, the raw data from one sensor in each array is transmitted to the fusion centre. The fusion centre then estimates the

propagation time delay between pairs of distributed arrays, and triangulates these time delay estimates with the bearing estimates to localize the source.

The first and second method can be realized easily by our hemispherical array. The bearing in 3D and a directional reference signal are two main outputs of our hemispherical array. Therefore, multiple array localization is easy to implement with eye array for large areas, especially if multiple upper corners are available.

7. Implementation Issues

This chapter deals with some implementation details and specifications in our research which should be mentioned to clarify this specific research. Hence it may not cover every aspect of our current system. We intentionally omitted the obvious and straightforward details. This chapter starts with a brief description of a preliminary test bench which helped us start working on the three-microphone arrangements prior to building the main array. Later, we discuss the mechanical aspects of building the array, followed by electrical features of the developed array system. Finally we discuss the implemented software and compare its computational efficiency with a cross correlation term.

7.1 Preliminary Test Bench

To test the pinhole cell and lens cell, we used a test bench as illustrated in Figure 7-1. The reference microphone and posterior microphone forms the pinhole cell and adding the lateral microphone fulfills the requirement for the basic lens cell.

Figure 7-1: Preliminary test bench

Instead of moving the sound source, a small loudspeaker, we rotated the sensor cell with the help of the turntable. The angles of rotation were measured simply by means of a protractor installed on turntable. We kept the distance from loudspeaker to microphones more than 2.5 meters to fulfill the far field assumption. Having such a test system, we could study the DCF cell, before building the main array. A sample normalized and filtered pinhole and lens DCF measured output versus angle shown in the following figure:

Figure 7-2 The pinhole (top) and lens (bottom) normalized measured DCF versus angle

7.2 Mechanical Design and Construction

Here we first go through some of the details of choosing the specified topology. Later we discuss the mechanical details for its implementation.

7.2.1 Topology Selection

In our algorithm, every posterior sensor has to have a corresponding lateral sensor. That is why we need a topology in which every vertex has another vertex or vertices with 90-degree angle difference between their axes from centre. This is not a general attribute of the majority of geodesic domes. In a two-frequency icosahedral geodesic dome, one can find plenty of orthogonal vertices for every other vertex. Figure 7-3 depicts the node numbers and angle difference of both types of spherical triangles.

Figure 7-3 Wireframe frontal view of array

Table 7-1: Eye array shell microphone directions with their orthogonal pairs

Direction i	Azimuth φ_i (degrees)	Elevation θ_i (degrees)	Orthogonal Directions $i^\perp$	P_i
1	0	0	17,18,19,20,21,22,23,24,25,26	10
2	0	31.72	11,12,13,19,24	5
3	30.45	9.35	13,14,15,21,26	5
4	19.96	-25.17	7,15,16,18,23	5
5	-19.96	-25.17	7,8,9,20,25	5
6	-30.45	9.35	9,10, 11,17,22	5
7	0	63.43	4,5, 10,14,19,24	6
8	43.56	43.49	5,11,15, 20, 25	5
9	62.27	16.05	5, 6,12,16,21,26	6
10	56.98	-15.24	6,7,13,17,22	5
11	49.61	-46.35	2,6,8,14,18,23	6
12	0	-58.28	2,9,15,19,24	5
13	-49.61	-46.35	2,3,10,16,20,25	6
14	-56.98	-15.24	3,7,11,21,26	5
15	-62.27	16.05	3,4,8,12,17,22	6
16	-43.56	43.49	4,9,13,18,23	5
17	90	72	1,6,10,15	4
18	90	36	1, 4, 11, 16	4
19	90	0	1,2,7,12	4
20	90	-36	1,5,8,13	4
21	90	-72	1,3,9,14	4
22	-90	-72	1,6,10,15	4
23	-90	-36	1,4,11,16	4
24	-90	0	1, 2,7,12	4
25	-90	36	1,5,8,13	4
26	-90	72	1,3,9,14	4

Table 7-1 shows the orthogonal pairs based on the above-mentioned numbering. The row i contains the vertex numbers which are orthogonal with the vertex number $i^\perp$.

As demonstrated, every vertex has at least four other orthogonal vertices in diverse locations and this is a surprising attribute of just this topology.

Furthermore, the array made by this topology, has 27 sensor nodes, which is a reasonable number regarding realization with computer add-on cards or integrated circuits.

The two-frequency icosahedral geodesic hemisphere subdivides the frontal view solid angle (2π steradians) into 40 spherical triangles. This topology is also known as Buckminster Fuller's Geodesic Dome [83]. It contains 10 equilateral triangles and 30 isosceles triangles. Groups of five of these isosceles triangles are used to replace each pentagon in the topology. Figure 7-4 illustrates a solid three dimensional representation of the array topology and its two dimensional net view. The blue triangles are isosceles and the red triangles are equilaterals.

Figure 7-4 Solid representation of the topology [94]

7.2.2 Array Design

The array structure was designed with combination of cylindrical rods connected to junction holders located at vertices (Figure 7-5). This modular structure was chosen to simplify any probable further change in the hemisphere size by merely changing the lengths of the rods. The lengths were chosen to form a hemispherical structure with a radius of 34 cm. Employing two protractors along with a mechanical rotation system, we could rotate the array in both azimuth and elevation angles.

Figure 7-5 Eye array design view

7.2.3 Mechanical Structure

A modular structure was chosen to simplify any possible further change in the hemisphere size by merely changing the lengths of the rods. The array structure was built with combination of cylindrical rods connected to junction holders located at vertices.

Figure 7-6 The constructed array structure

Both rods and joints were made by a milky polyethylene plastic. The simplicity of machining plastic joints persuaded us to use this material, but after a while the plasticity of rods caused the array to bow down owing to its weight and create error due to misplacement. The joints in this topology are two types of five sided and six sided (Figure 7-6, Figure 7-7). The joint carrying the reference microphone is three sided and connected to the array corners with three symmetrical rods.

Figure 7-7 Microphone joint holders

A special mechanical rotation system for azimuth and elevation was built. Two protractors with swinging arm indicators create goniometers for measurement of the array direction (Figure 7-8). Both of these rotational systems rotate the array from the bottom corner of the array instead of the array center. One can easily prove that, if we fix the array central direction at $\theta = \varphi = 0$, after all further rotations the final azimuth and elevation read by goniometers at bottom corner is exactly equal to real azimuth and elevation of center of array, although the array center has moved away from its initial position due to rotation from the bottom. In mathematical words, the rotational difference remains constant if we move the center of rotation from center of the hemisphere to anywhere on its shell although each case ends up in different array positions.

Therefore having the rotation and measurement system on the corner of hemispherical array, does not apply any extra transformation to the measured azimuth and elevation difference.

Figure 7-8 Array goniometers

To verify the abovementioned statement, consider we have a coordinate in the center of a hemisphere which is movable via rotation of a fixed coordinate system located at bottom corner of hemisphere. Both coordinates are initially aligned but the movable coordinate can freely rotate around the fix coordinate within a fixed distance r . Moreover consider a constant point P in space measurable by the movable coordinate.

Figure 7-9 Fixed coordinate at corner versus movable coordinate at center of hemisphere

To map the movable coordinate (x_0, y_0, z_0) to the fixed coordinate (x, y, z) , we have to resolve each of the unit vectors of the fixed coordinate to the unit vectors of the movable coordinates. For example i_x can be resolved into its components, via the supplementary coordinate (x', y', z) . The y' is along the intersection of the $\theta = \text{constant}$ and $x - y$ plane and x' is perpendicular to the $y' - z$ plane. As a result, $\sin(\theta)$, $\cos(\theta)$ and 0 are the components of i_x along the x', y' and z axes respectively. These components are in turn resolved into components along the movable coordinate (x_0, y_0, z_0) directions by recognizing that the component $\sin(\theta)$ along the

x' axis is in the $-y_0$ direction, while the component of $\cos(\theta)$ along the y' axis resolves into components $\cos(\theta)\cos(\varphi)$ in the direction of x_0 , and $\cos(\theta)\sin(\varphi)$ in the z_0 direction. Therefore:

$$i_x = \cos(\theta)\cos(\varphi)i_{x_0} - \sin(\theta)i_{y_0} + \cos(\theta)\sin(\varphi)i_{z_0} \quad (7.1)$$

Similarly,

$$i_y = \sin(\theta)\cos(\varphi)i_{x_0} + \cos(\theta)i_{y_0} + \sin(\theta)\sin(\varphi)i_{z_0} \quad (7.2)$$

$$i_z = -\sin(\theta)i_{x_0} + \cos(\varphi)i_{z_0} \quad (7.3)$$

The coordinate of a fixed point P in the fixed coordinate (x, y, z) , is the sum of the coordinate of P in (x_0, y_0, z_0) mapped into (x, y, z) plus the vector $\vec{r}$ linking the two coordinates:

$$\vec{R} = \begin{pmatrix} \cos(\theta)\cos(\varphi) & -\sin(\theta) & \cos(\theta)\sin(\varphi) \\ \sin(\theta)\cos(\varphi) & \cos(\theta) & \sin(\theta)\sin(\varphi) \\ -\sin(\theta) & 0 & \cos(\varphi) \end{pmatrix} \vec{r}_0 + \vec{r} \quad (7.4)$$

Where:

$$\vec{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = r \begin{bmatrix} \cos(\theta)\sin(\varphi) \\ \sin(\theta)\sin(\varphi) \\ \cos(\varphi) \end{bmatrix} \& \quad \vec{r}_0 = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = r_0 \begin{bmatrix} \cos(\theta_0)\sin(\varphi_0) \\ \sin(\theta_0)\sin(\varphi_0) \\ \cos(\varphi_0) \end{bmatrix} \quad (7.5)$$

Although (x_0, y_0, z_0) is movable, $\vec{R}$ is the coordinate of the fixed point P in the fixed coordinate (x, y, z) ; therefore:

$$d\vec{R} = 0 \quad (7.6)$$

Taking the derivative of each of the pairs in (7.5) leads us to:

$$\begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} = \begin{pmatrix} \cos(\theta)\sin(\varphi) & -r\sin(\theta)\sin(\varphi) & r\cos(\theta)\cos(\varphi) \\ \sin(\theta)\sin(\varphi) & r\cos(\theta)\sin(\varphi) & r\sin(\theta)\cos(\varphi) \\ \cos(\varphi) & 0 & -r\sin(\varphi) \end{pmatrix} \begin{bmatrix} dr \\ d\theta \\ d\varphi \end{bmatrix} \quad (7.7)$$

By taking the derivative of (7.4) while using (7.5) and (7.7) also carefully following the huge and tedious job of simplifying both sides while considering $dr = 0$, we finally end up:

$$\begin{aligned} d\theta &= d\theta_0 \\ d\varphi &= d\varphi_0 \end{aligned} \quad (7.8)$$

In other words, the changes of θ and φ are equally reflecting to the changes of θ_0 and φ_0 .

7.3 Electrical Implementation

Here we go through the electrical aspects of our laboratory implemented eye array. Microphones, the data acquisition board, preamplifiers, wirings, sound source and the computer will be discussed briefly. Later we describe the software along with a benchmark test. Finally we suggest an integrated circuit implementation for utilization of the algorithm.

7.3.1 Microphones

The selection of microphones for our work is based on three aspects, omnidirectionality, small size, and low cost. There are different Electret condenser microphones with these features in market. Most of them come with two or three wires and have EMI and RFI shielding for minimum induced noise. They are cylindrical shape and as small as 6mm x 2.2mm in size and quite inexpensive in larger quantities. Some of

the characteristics of a typical Panasonic microphone of this type have been illustrated in Figure 7-9 [95, 96].

Figure 7-10 Microphone shape, directivity pattern, connection and frequency response

During the course of our research we initially used a Panasonic Electret microphone, but later due to need for lower induced noise and higher signal to noise ratio we switched to an omni-directional Electret microphone with built in integrated circuit preamplifier, made by Best Sound Electronics (BSE) Model No: BGO-15L27-C1033.

This microphone has a smaller size (cylinder of 6mm diameter and 1.5mm height) compared with the previous one (8mm by 2.5mm). Therefore to keep the new microphone attached to their previous location in the junction nodes as well keeping

them exactly on the shell of hemisphere, we made 27 extra plastic holders to fit the new microphones in place of the old ones.

The suggested measurement circuit of this microphone is shown in Figure 7-10.

Figure 7-11 Measurement circuit for the BSE microphone

Instead of using the suggested RC circuit in Figure 7-11, we used the pre-filter stage of our preamplifier/filter board as a low impedance current to voltage converter and isolator. This reduced the amount of induced noise in microphone wires by reducing the termination impedance.

The far field frequency response of the BSE microphone is shown in Figure 7-11.

Figure 7-12 Frequency response of BSE microphone

As you can see from Figure 7-11 the frequency response of this microphone rises after 10 kHz due to the overshoot response of its preamplifier (National® LMV1024). Variable gains of several samples of this microphone showed that this overshoot is not even consistent among all samples. Therefore we decided to limit the maximum frequency of our tests with this microphone to 8 kHz, in order to stay in the flat range of its frequency response.

7.3.2 Data Acquisition Board

We acquired all of the 27 single ended voltage outputs of the array by utilizing 27 channels of a 64 channel National Instruments® data acquisition PCI board “NI-6071E” (Figure7-13) [97]. This is a multifunction data acquisition add-on board with a sum of 1.25 MS/s sampling rate for all channels, 12-Bit A/D resolution, 32 Differential /64 Single Ended analog voltage inputs. It also contains 2 analog outputs, 8 digital I/O lines, two 24-bit counters as well as an analog triggering capability.

Figure 7-13 National Instrument PCI-6071E data acquisition board

In addition, we used a noise rejecting shielded I/O connector block (SCB-100) along with a shielded 2 meter SH-100-100F connector cable (Figure 7-13) to connect the data acquisition board to the connector block.

Figure 7-14 Cable and connector block

7.3.3 Preamplifiers

Initially we connected the microphones directly to the DAQ card for our tests, but later due to several reasons we decided to add a preamplifier/ filter stage between them. The additional preamplifier, notch and low pass filter was designed to filter unwanted line and high frequency components as well enhancing overall amplification.

Although the microphones which we used have a 6 dB built in amplifier IC inside them, we needed extra amplification and pre-filtering to reach a reasonable level of signal at the input stage of the DAQ board. As well, to reduce the induced noise in long microphone wires (average 2.5 meters), we preferred to terminate them to a low impedance current to voltage converter instead of using simple resistors that are recommended by Figure 7-10. Furthermore to reduce the unpleasant effect of ghosting in our DAQ board, we forced to supply the input of the data acquisition board with low impedance.

Figure 7-15 Preamplifier/Filter daughter boards

Hoth noise is used to model indoor ambient noise when evaluating communications systems. It is named after D.F. Hoth, who made the first systematic study of the ambient noise [98]. Based on the frequency content of Hoth noise, most of the energy of acoustic noise lies on lower frequencies. The fact that the line frequency and its third harmonic also stays in microphone signal urged us to high pass filter the signal after 200Hz. Since the power of speech signal dramatically decreases after 8 kHz, we added a low pass filter at 8 kHz as well. Both lowpass and highpass filters are designed and combined with the gain stage.

7.3.4 Computer

The computer is a desktop with Pentium IV CPU operating at 2.4GHz and 512Kbytes RAM. It has an onboard sound card that was employed to drive the loudspeaker. The loudspeaker is a small desktop computer loudspeaker installed on a vertically adjustable

stand. A laser pointer is attached to its front surface to assist the speaker-array direction initial setting task (Figure 7-16).

Figure 7-16 The Loudspeaker on stand with the attached laser pointer

For noise measurements we utilized an integrating sound level meter made up of Bruel & Kjaer® Type 2225 (Figure 7-17). Although it applies a special prefiltering called A-weighting to the signal to compensate for perceived loudness, in our frequency range (0.3-4 kHz) the filter does not have much effect [85] but creates an accurate referenced omni-directional output suitable for SNR measurement purpose.

Figure 7-17 Brüel & Kjær integrating sound level meter

7.3.5 Software

Initially we used Matlab® (v.7.0) Data Acquisition Toolbox, but during maximum data rate acquisition with NI®-PCI-6071E some of the channels experienced data loss causing erroneous delays among channels. Therefore we switched to LabView® (v.7.1) for both data acquisition and algorithm implementation.

The LabView® experiment platform for MCF is shown in Figure 7-18. As illustrated, in addition to the main signal path, there are extra sections for visualization, initial array adjustment, sound generation and writing.

Figure 7-18 The MCF experiment platform in LabView

Figure 7-19 shows only the signal path of the MCF algorithm. There are two subdivisions in the program. The first deals with the calculation of lens closeness functions and the second deals with the sorting the closeness functions, choosing k maximum closeness function and the final estimation.

Figure 7-19 The MCF algorithm

7.3.6 Computational Complexity

The parallel mathematical calculation of all closeness functions is the key to a low computational load for sound source localization. Here we try to show the lower computational complexity of the eye array by first analyzing the computation burden and later a benchmarking comparison of the eye array versus a convolution pair.

In the eye array algorithm there are two distinct sections. The first is the calculation of the closeness functions. The latter is the part dealing with finding the k maximum closeness functions and the linear spatial averaging. This section does not have

demanding computations. The integration time or data frame number, has no effect on the second part, but influences the closeness function calculation. Therefore the calculation of closeness functions is the major computational part in the eye array algorithm.

Consider the MCF in (3.22), we need only to calculate the term $\overline{\langle [S_0(t) - S_i(t)].[S_0(t) - S_0(t - T_r)] \rangle}$ once for all directions for a specified length of N data samples. This needs N multiplications, $26(N+1)$ subtractions, $26N$ additions and N delay operations. Finally we have to include 26 divisions for the MCF in (3.22). The computational cost of multiplication and division (multiplicative operations) is so higher than addition and subtraction (additive operations). Thus, generally the number of multiplicative operations resembles the complexity of the algorithm. To calculate all 26 MCFs, we need $26N+26$ multiply and $26(2N+1)$ add operations for the sound signals with length N .

The foremost part of all TDOA based sound source localization methods is the calculation of multiple two-channel cross correlations (Section 2.2). Published research [1, 2, 99, 100] utilizes more than 4 pairs of microphones for sound source localization in 2D (localization within a plane) applications.

Therefore we claim that the computational burden of one pair of cross-correlations can be considered as a suitable reference for a complexity comparison of our algorithm.

The cross correlation of a pair of signals with length N needs N^2 real multiplies if performed in the time domain. By benefiting from the computational speed of the Fast Fourier Transform (FFT) in longer samples restricted to lengths $N = 2^n$, one can calculate the modified and simplified version of (2.14):

$$CPS_k(n) = IFFT[FFT\{S_i(n)\}.FFT\{S_k(n)\}^*] \quad (7.9)$$

The Cooley-Tukey [101] FFT algorithm, needs $N \log_2(N)$ complex multiplications ($2N \log_2(N)$ real multiplications) for each length N FFT. Even if we ignore the length extension factor and zero padding required to prevent overlap, performing (7.9) needs $6N \log_2(N)$ real multiplication for two FFTs and one IFFT and N real multiplies for the multiplication of two FFTs. By comparing the number of multiplications for (7.9) ($6N \log_2(N) + N$) and the number of multiplications for all MCFs ($26N+26$), one can figure out that for lengths $N \leq 16$ the computation cost of (7.9) is higher and for lengths greater than $N=16$ the computational cost of (7.9) is higher than MCF calculation. The difference rises with a $\log_2(N)$ ratio. Similar behavior holds if one compares the number of additive operations in (7.9) compared with 26 MCFs.

FFT algorithms have enhanced since the basic Cooley-Tukey version. The maximum achieved for modulo 2 lengths does not exceed more than 20% enhancement in 1-D FFT [102]. Even after enhancements, the computational complexity is still $N \log_2(N)$ dependent.

To evaluate the computational complexity of the eye array algorithm, we set up a benchmark to numerically compare the performance times of the closeness function calculation section, with the simple two-channel cross-correlation in (7.9). Figure 7-20 (top) shows the benchmark test program for the two-channel cross correlation. We chose the most efficient block in LabView® to perform the cross-correlation. This block performs based on taking the FFT of each of two signals, multiplying them and taking the IFFT in a single efficient program.

Also Figure 7-20 (bottom) shows the benchmark test for calculation of 26 pinhole and lens closeness functions.

Both of the benchmarking programs consist of a three frame flat sequences. All of main code is performed inside the middle frame. A tick counts the start of the first frame and third frame. The subtraction of the first tick and the third tick is almost the middle frame's execution time. To have a more reasonable time estimate by averaging, we put the code inside a loop in the second frame. To narrow down the effect of variable resources such as any probable memory change or other running programs effect, we tested each of the benchmarks after a system restart without any other programs opened.

Figure 7-20 Computational performance comparison benchmarks

For a loop of 1000 runs with dummy input signals of the length 2048 sample, cross correlation shows the total time of 840 milliseconds and the multiplicative closeness functions takes 273 milliseconds. In other words, one pair of cross correlations takes approximately 0.84 millisecond and 26 closeness functions takes 0.27 milliseconds. By increasing the frame length to 4096, 8192 and 16384; the cross correlation time is 2014, 4467 and 9625 respectively, while the closeness function times are 511, 986 and 2098 milliseconds for 1000 repetitions. This tells us that the whole 26 channel closeness function calculation is 3 to 5 times faster than cross correlation. As stated cross correlation computation load increases with $N \log_2 N$ while closeness function computation load increases with N .

Mathematically speaking, if one subdivides a vector into N parts, the average of the vector is equal to the average of all N part averages. This property also holds for average of the inner product of two vectors.

All endfire and broadside closeness functions that we have described so far, consists of either the average of vectors {DCF (4.11) and (4.12)} or inner product of two vectors {MCF (3.11),CCF (4.28)}. Therefore to calculate a long endfire or broadside term, one can subdivide the data into shorter lengths and take the average of all parts. The blocked moving average (BMA) in the middle part of our program is written for this purpose.

For an efficient closeness function digital implementation, instead of having a long frame length, we can keep sampling with minimum block length, and add up the smaller frame length results for each endfire and broadside term. In a blocked moving average, similar to a moving average, increasing the number of blocks does not increase the

computational complexity. Therefore we can calculate a 16384 length frame closeness function by keeping the frame length to 2048 and set the BMA counter to 8.

The closeness function execution time for an initial frame length of 2048 and BMA of 50 is 292 milliseconds, which is a 7% increase compared with single 2048 length frame (273 msec).

Clearly if we subdivide a vector and average its cross correlations, the result is not the average of cross correlation of parts.

The computational advantage of the eye array over TDOA based methods becomes more evident in longer time frames. The simple causal and time domain computation encourages us to consider analog implementation as well.

7.4 Method of Choice: Integrated Circuit

Most of the sound source localization methods utilize a computer with add-on boards or specifically designed DSP boards. Rarely any of them, as a whole have the capability of being implemented on a chip, mainly because of their high computational complexity.

There are some microphone array related researches, which utilize integrated circuits for beam forming in a specified direction [103]. Others utilize analog chips for computing time delays between sound signals for inter-aural time delay cues as part of a localization system [104, 105].

The group of researchers in The Artificial Perception Lab® in University of Toronto precisely addressed the need for an integrated circuit implementation of sound source localization of one pair of microphones [106], for probable applications in hand-held devices and limited battery life situations. They designed a single field programmable

gate array (FPGA) implementation of a real-time sound localization system using two microphones. The implementation utilizes a cross-correlation technique based on a modified version of GCC-PHAT. Later they changed their implementation from the FPGA to a custom designed digital integrated circuit with a $0.18\mu m$ CMOS process [107]. They also designed a dual-microphone phase-based speech enhancement FPGA. By using the phases of the incoming sound signals, they can mask low SNR frequencies between microphone pairs [108]. The FPGA implementation was also compared with an off-the-shelf digital signal processor (DSP) implementation with respect to processing capabilities and power utilization [109].

The eye sound source localization method benefits from some simple, parallel routines at the closeness function stage as well as a simple maximum/minimum decision plus linear estimation at the estimation stage, it can be implemented on a mixed-signal or analog integrated circuit. This obviously reduces the manufacturing cost of the probable product astonishingly as well as enhances its identical performance versus all angles. Moreover in most products and applications the implementation ought to be cheap, lightweight and portable, to be known as ubiquitous and affordable. For all these reasons, an integrated circuit (IC) implementation is preferred. This IC has to exploit 27 input channels for 27 microphone inputs and two output channels for azimuth and elevation and possibly another output for a beam formed sound signal. It would be possible to send the data wirelessly to a central station or a host computer.

An evident drawback of our array is its relatively high number of microphones. Each microphone has to be sampled with analog to digital (A/D) channels before any processing. This increases the cost of implementation in any digital realizations. The

difficulty of having 27 A/D in an IC is obvious. The analog implementation will take out the need for A/D channels. Moreover, an analog implementation can benefit from the lack of A/D quantization noise, despite the fact that it introduces other types of noises. The algorithm consists of some simple addition, multiplication and division operations. These calculations are similar in all cells. Therefore, the final design can be repeated for all channels. Moreover, the analog processing is the same as sampling with infinite sampling rate and therefore keeps us in maximum achievable resolution.

It is well known that nonlinear systems may produce output signals with larger bandwidth due to spectral spreading. Working in analog domain preserves the ability of higher frequency signals and removes the requirement of higher sampling rates.

8. Conclusions

In this final chapter of the thesis, achievements are discussed first, followed by the disadvantages of this implementation, and a list of possible future research topics.

8.1 Summary of Thesis Contributions

This research has shown that our localization approach, the eye array, works reliably and yields the expected results in the localization of the sound source. A general framework based on a hemispherical array and closeness functions has been presented. Three different categories of closeness functions have been introduced. The localization coverage and accuracy is reasonable and the computation cost is remarkably low in comparison to other methods.

Figure 8-1 Theoretical (left) and experimental (right) coverage

Although the method promises 2π steradians ($\pm 90^\circ$ azimuth and elevation) theoretical coverage, we obtained $\frac{4}{3}\pi$ steradians ($\pm 60^\circ$ azimuth and elevation) three dimensional coverage (Figure 8-1). In the experimental setup, MCF processing yielded an accuracy of 3.1 degrees and a precision of 0.69 degrees (defined in Section 5.1).

The significant achievements of the eye array are a) 3D direction coverage and b) very low computational cost due to its symmetrical geometry. As discussed in Section 7.3.6, for 1000 loops of data with a frame length of 2048 samples the TDOA algorithm takes 840 msec. processing time while MCF takes 273 msec. processing time. By increasing the frame length to 16384 samples, the TDOA processing time increases to 9625 msec., while the MCF takes 2098 msec. of processing time. So, the closeness function calculation which is roughly the main computational part of eye array algorithm, in our specific implementation, is more than three to five times faster than the computation of a simple two-channel cross correlation which is the building block of the pairwise TDOA algorithm (Section 2.2), even without using a moving average.

Considering the high probability of error in pairwise TDOA and high rate of false peak detection due to reverberation, in a reasonable system one would have to utilize at least four pairs of microphones. Thus we can claim that our system is at least 12 times faster than an 8 microphone TDOA based sound source locator. It might have better precision and accuracy if implemented in a chip.

A novel placement strategy for enclosed areas has also been offered. This special placement benefits the retroreflection property of trihedral corners, low sensitivity of

closeness functions to back reverberation and symmetry to reduce the adverse effect of early reverberation, on the performance of eye array system.

A number of publications have been written based on this research to date. [110, 111,112,113,114,115 116,117,118,119]

8.2 Disadvantages

The eye array SSL implementation that we have described has some drawbacks. Some of them are common among other methods. Here we categorize the general drawbacks of our approach:

- **Array size:** The manufactured array is a hemisphere with a 34cm radius. Although it is smaller than most of its counterparts, it still covers a significant portion of a room.
- **High number of microphone channels:** This array has 27 signal channels. In order to implement the system on a computer, the need for a data acquisition board with 27 analog to digital channels is problematic. Therefore a dedicated integrated circuit implementation is suggested (Section 7.4).
- **Array gain adjustment:** It seems that the difference cell is more sensitive to calibration error than the multiplicative and correlative cells. This forces us to add a gain adjustment mechanism for the proposed method, if a DCF is implemented.
- **Temperature dependence:** Since the speed of the sound in a medium varies with the temperature of the medium, we have to adjust our delay based on the room

temperature. This can easily be done by installing a small temperature sensor, e.g., located back to back with the reference microphone and adjusting the delay $T_r = r / c$ based on the room temperature.

- Sensor position error: This is a general drawback of every sensor array system. Arrays that have a connected array topology have lower error possibility compared to those that consist of multiple array structures, which are not physically connected [1]. Since in our proposed method, the microphones are attached to a connected solid structure (geodesic dome), one can anticipate lower position errors compared with TDE based methods that utilize separated arrays of mostly two-microphones in the room.

8.3 Future Directions and Enhancements

Here we first explain our perception of current research in sound source localization and capturing and its probable future. Later we discuss future possibilities and enhancements in sound source localization in general and the eye array system in particular, for the sake of potential forthcoming researchers in this area.

8.3.1 Perspective

Remotely capturing and localizing of sound sources via acoustic cues is not a popular current research topic at this period of time. There was a wave of microphone array related research starting in the early nineties, due to the low cost of microphones and the

advent of computerized signal processing systems¹[120]. Immediate achievements were gained and obstacles were recognized. There has not been much major new academic research on this topic since. Most of the newer work, if any, deals with enhancing previous methods to some extent by researchers in the universities that had a history in this area.

Although in our opinion the trend of SSL related research was declining, the future prospect for this area may recover in the long run. Technology visionaries deem that human sound will play greater roles in control of computer systems (operating systems) and high technology devices in future. This happens with the advent of highly robust speech recognition and processing systems. We believe that, if this prevails, due to a human desire to navigate freely around a room while speaking or working, the need for remote sound capture and localization will increase subsequently.

8.3.2 Approach

Sound localization and capture used to be thought of as a sole system and algorithm. We believe that this concept is not valid anymore. Sound localization does not need to utilize a high sensitivity, expensive single microphone or arrays of high end microphones. Furthermore, scanned beamforming of a whole line, plane or space (focalization) requires huge computational resources and is not a solution affordable for everyone or every application.

¹ SSL research started by researchers in four major academic labs (Rutgers Univ., Brown University, Harvard University and IRST in Italy). Most of the researchers involved in SSL no longer work in SSL or SSL is not their primary research topic. The aforementioned academic labs either no longer are active or shift their primary research from SSL to speech processing related topics. Furthermore, speech processing researches is the main topic of recent sound related academic labs. Almost all current SSL research are singleton work.

Therefore we consider that in an ultimate solution, the sound localizing system is separate from the sound capturing system. This not only reduces the cost of equipment, but also reduces the computational cost by allowing the use of parallel non-linear calculations that possibly will give us information about the sound direction, intensity or frequency, without any exact ability to retrieve the sound signal.

Fusing of geometrical localization of sound sources, we suggest some new research directions that can enhance eye array sound source localization, such as utilizing a smaller array size and better physical implementation and structure which provides higher omni-directionality for all microphones on the array body. Integrated circuit (analog-mixed signal) implementation, instead of digital software definitely increases the resolvability and overall performance of the system.

We also believe that the whole notion of estimation of sound sources in a complex sound field with a limited number of sensors in the space is arguable.

Having increased the number of microphones on the shell while increasing their directivity; tends to build images of the sound field viewed from virtually a point. Increasing the number of channels, by increasing the detection ability of each cell leads to the formation of real time images of the sound. Thus by using the vast resources of vision and image processing on the sound image, one can detect multiple human sound sources (talkers) out of other sound sources and reverberated shadows. The sound image has to be generated without tedious and huge processing like array beamforming and focalization, to represent a low-cost alternative sound source localization to the current methods. Using unconventional materials and building acoustic lenses could be an alternative way to beamform sound and utilize a sound camera.

Finally, in sound source localization one should not just rely on the free field sound capture with arrays of microphone. Utilizing cavities can extend the time delays and change directions to a great extent. As well using the recent paradigms in vibration evaluation like optical laser measurements may take away the need of using ordinary microphones as the only medium for capturing the behaviour of the sound fields.

References

- [1] Brandstein M., Ward D., Microphone arrays: signal processing techniques and applications, New York, Springer, 2001.
- [2] Gay S. L., Benesty J., Acoustic signal processing for telecommunication, Kluwer Academic, Boston, 2000.
- [3] Brandstein M., Silverman H. F., "A practical methodology for speech source localization with microphone arrays," Computer, Speech, Language, vol. 2, pp. 91-126, Nov. 1997.
- [4] Brandstein M., "A Framework for Speech Source Localization Using Sensor Arrays", PhD thesis, Brown University, Providence, RI, May 1995.
- [5] Wax M., Kailath T., "Optimum localization of multiple sources by passive arrays", IEEE Trans. Acoustic, Speech, Signal Processing, vol. 31, pp. 1210-1217, October 1983.
- [6] Carter G., "Variance bounds for passively locating an acoustic source with asymmetric line array," J. Acoust. Soc. Am., vol.62, pp. 922-926, October 1977.
- [7] Hahn W., Tretter S., "Optimum processing for delay-vector estimation in passive signal arrays," IEEE Trans. Inform Theory, vol.19, pp. 608-614, September 1973.
- [8] Egemen Gonen, Jerry Mendel, "Subspace Based Direction Finding Methods" in "DSP Handbook", CRC Press, 1999.
- [9] Simon Haykin, Adaptive Filter Theory, Prentice Hall, second ed., 1991.
- [10] Wax M., Kailath T., "Optimum localization of multiple sources by passive arrays," IEEE Tran. Acoust. Speech, Signal Processing, vol. ASSP-31, pp. 1210-1217, October 1983.

-
- [11] Wang H., Kaveh M., "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," IEEE Trans. Acoust, Speech, Signal Processing, vol. ASSP-33, pp. 823-831, August 1985.
- [12] Shan T., Wax M., Kailath T., "On spatial smoothing for direction-of arrival estimation in coherent signals," IEEE Tran. Acoust, Speech, Signal Processing, vol. ASSP-33, pp. 806-811, August 1985.
- [13] Svaizer P., Matassoni M., Omologo M., "Acoustic source location in a three-dimensional space using cross-power spectrum phase," in Proc. IEEE Int. Conf. Acoust. , Speech, Signal Processing (ICASSP-97), Munich, Germany, pp. 231- 234, April 1997.
- [14] Omologo M., Svaizer P. Svaizer. "Acoustic event localization using a cross power-spectrum phase based technique." Proceedings of ICASSP-1994, Adelaide, Australia, 1994.
- [15] Omologo M., Svaizer P., "Acoustic source location in noisy and reverberant environment using CSP analysis", Acoustics, Speech, and Signal Processing, ICASSP-96. Conference Proceedings, IEEE International Conference on , Atlanta, GA, May 1996.
- [16] Omologo M., Svaizer P., "Use of the crosspower-spectrum phase in acoustic event location", Speech and Audio Processing, IEEE Transactions on, Vol. 5 , Issue: 3 , pages: 288 – 292, May 1997.
- [17] Piersol A., "Time delay estimation using phase data", Acoustics, Speech, and Signal Processing, IEEE Transactions on Signal Processing, Vol. 29, Issue: 3, pages: 471- 477, Jun 1981.
- [18] Knapp C., Carter G., "The generalized correlation method for estimation of time delay." IEEE Transactions on Acoustics Speech and Signal Processing. Vol.24, no.4, August 1976.

-
- [19] Daniel V. Rabinkin, "Optimum sensor placement for microphone arrays", Ph.D. Thesis, Rutgers University, May 1998.
- [20] Brandstein M., "A Framework for Speech Source Localization Using Sensor Arrays", PhD thesis, Brown University, Providence, RI, May 1995.
- [21] Aarabi P., Mahdavi A., "The relation between speech segment selectivity and source localization accuracy", Acoustics, Speech, and Signal Processing, Proceedings (ICASSP '02). IEEE International Conference on , Vol.1, 2002.
- [22] Brandstein M., Adcock J., Silverman H., "A closed-form location estimator for use with room environment microphone arrays," IEEE Trans. Speech Audio Proc., vol. 5, pp. 45-50, January 1997.
- [23] Chan Y.T., Ho K.C., "A simple and efficient estimator for hyperbolic location", Signal Processing, Acoustics, Speech, and Signal Processing, IEEE Transactions on, Vol. 42 , Issue: 8, Pages: 1905 – 1915, Aug. 1994.
- [24] Gillette M. D., Silverman H. F., "A Linear Closed-Form Algorithm for Source Localization From Time-Differences of Arrival", Signal Processing Letters, IEEE, Vol.15, Pages:1-4, January 2008.
- [25] Vahedian A., Frater M., Arnold J., Cavenor M., Godara L., Pickering M., "Estimation of speaker position using audio information", TENCON '97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications, Proceedings of IEEE, Volume 1, Pages:181 – 184, Dec. 1997.
- [26] Ying Yu, Silverman H.F., "An improved TDOA-based location estimation algorithm for large aperture microphone arrays", Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on, May 2004.

-
- [27] Griebel S., Brandstein M., "Microphone Array Source Localization Using Realizable Delay Vectors," IEEE Workshop on Applications of Signal Processing To Audio and Acoustics, New Paltz, NY, October, 2001.
- [28] Griebel S., "Multi-Channel Wavelet Techniques for Reverberant Speech Analysis and Enhancement," Harvard Intelligent Multi-Media Environment Laboratory Technical Report, February 1999.
- [29] Parham Aarabi, "Spatial Integration and Localization of Dynamic Sensors", Ph.D. Thesis, Stanford University, May 2001.
- [30] Jahromi O., Aarabi P., "Distributed spectrum estimation in sensor networks", Acoustics, Speech, and Signal Processing, 2004. Proceedings (ICASSP '04). IEEE International Conference on , May 2004.
- [31] Aarabi P., "The Fusion of Distributed Microphone Arrays for Sound Localization", EURASIP Journal of Applied Signal Processing (Special Issue on Sensor Networks), Vol. 2003, No. 4, pp. 338-347, March 2003.
- [32] Brutti A., Omologo M., Svaizer P., Zieger C., "Classification of Acoustic Maps to Determine Speaker Position and Orientation from a Distributed Microphone Network", Acoustics, Speech and Signal Processing, 2007 ICASSP 2007, IEEE International Conference on , page: IV-493-496, Honolulu, HI, April 2007.
- [33] Macho D., Padrell J., Abad A., Nadeu C., Hernando J., McDonough J., Wolfel M., Klee U., Omologo M., Brutti A., Svaizer P., Potamianos G., Chu S.M., "Automatic Speech Activity Detection, Source Localization, and Speech Recognition on the CHIL Seminar Corpus", Multimedia and Expo, ICME 2005, IEEE International Conference on, Amsterdam, July 2005.
- [34] Xiaohong Sheng, Yu-Hen Hu, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks", Signal Processing,

IEEE Transactions on Acoustics, Speech, and Signal Processing, IEEE Transactions on, Vol. 53, Issue: 1, pages: 44- 53, Jan. 2005.

[35] Ajdler T., Kozintsev I., Lienhart R., Vetterli M., "Acoustic source localization in distributed sensor networks", Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on, Nov. 2004.

[36] Jahromi O.S., Aarabi P., "Time delay estimation and signal reconstruction using multi-rate measurements", Acoustics, Speech, and Signal Processing, 2003. Proceedings (ICASSP '03) 2003 IEEE International Conference on, Vol. 6, April 2003.

[37] Jahromi O.S., Aarabi P., "Theory and design of multirate sensor arrays", Signal Processing, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 53, Issue: 5, pages1739–1753, May 2005.

[38] Handzel A. A., Krishnaprasad P. S., "Biomimetic Sound-Source Localization," IEEE Sensors Journal, vol. 2, pp. 607-616, December 2002.

[39] Sturim D.E., Brandstein M.S., Silverman H.F. , "Tracking multiple talkers using microphone-array measurements", Acoustics, Speech, and Signal Processing, IEEE International Conference on, ICASSP-97., 1997.

[40] Douglas E. Sturim, "Tracking and Characterizing of Talkers Using a Speech Processing System with a Microphone Array as Input", Ph.D. Thesis, Brown University, 1999.

[41] Rafaely B., "Plane-wave decomposition of the sound field on a sphere by spherical convolution," Journal of Acoustic Society of America., Vol.116, pages: 2149–2157, 2004.

[42] Park M., Rafaely B., "Sound field analysis by plane wave decomposition using spherical microphone array", Journal of Acoustic Society of America., Vol.118, pages: 3094–3103, 2005.

-
- [43] Mungamuru B., Aarabi P., "Joint sound localization and orientation estimation", Information Fusion, 2003. Proceedings of the Sixth International Conference of , Vol. 1, pages: 81 - 85 , 2003.
- [44] Mungamuru B., Aarabi P., "Enhanced sound localization", Systems, Man, and Cybernetics, Part B, IEEE Transactions on, Vol. 34, Issue: 3, pages: 1526- 1540, June 2004
- [45] Aarabi P., Mungamuru B., "Scene reconstruction using distributed microphone arrays", Multimedia and Expo, ICME '03. Proceedings 2003 International Conference on , page: III - 53-6 vol.3 , July 2003.
- [46] Bob Mungamuru, "Enhanced sound localization", M.Sc. Thesis, , University of Toronto, 2003.
- [47] Sachar J.M., Silverman H.F., "A baseline algorithm for estimating talker orientation using acoustical data from a large-aperture microphone array", Acoustics, Speech, and Signal Processing, 2004. Proceedings (ICASSP '04). IEEE International Conference on, May 2004.
- [48] Brutti A., Omologo M., Svaizer P., "Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays", Interspeech, Lisbon, Portugal, September 2005.
- [49] Brutti A., Omologo M., Svaizer P., "Speaker Localization based on Oriented Global Coherence Field", Interspeech, Pittsburgh, PA, USA, September 2006.
- [50] Parham Aarabi, "Self-localizing dynamic microphone arrays", Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, Vol. 32, Issue: 4, pages 474-484, Nov. 2002.

-
- [51] Ward D.B., Lehmann E.A., Williamson R.C., "Particle filtering algorithms for tracking an acoustic source in a reverberant environment", *Speech and Audio Processing, IEEE Transactions on*, Vol. 11, Issue: 6, pages: 826- 836, Nov. 2003.
- [52] Ward D.B., Williamson R.C., "Particle filter beamforming for acoustic source localization in a reverberant environment", *Acoustics, Speech, and Signal Processing, 2002. Proceedings (ICASSP '02) IEEE International Conference on*, Orlando, FL, pages: 1777-1780, May 2002.
- [53] Parisi R., Croene P., Uncini A., "Particle swarm localization of acoustic sources in the presence of reverberation", *Circuits and Systems, 2006. ISCAS 2006 Proceedings 2006 IEEE International Symposium on*, May 2006.
- [54] Huang Do, Silverman H.F., Ying Yu, "A Real-Time SRP-PHAT Source Location Implementation using Stochastic Region Contraction(SRC) on a Large-Aperture Microphone Array", *Acoustics, Speech and Signal Processing, 2007, ICASSP 2007, IEEE International Conference on*, April 2007.
- [55] Duraiswami R., Zotkin D., Davis L.S., "Active speech source localization by a dual coarse-to-fine search", *Acoustics, Speech, and Signal Processing, 2001. Proceedings (ICASSP '01). 2001 IEEE International Conference on*, Salt Lake City, UT, USA, 2001.
- [56] Zotkin D.N., Duraiswami R., "Accelerated speech source localization via a hierarchical search of steered response power", *Speech and Audio Processing, IEEE Transactions on*, Vol. 12, Issue: 5, Sept. 2004.
- [57] Li Z., , Duraiswami R., "Fast Time-Domain Spherical Microphone Array Beamforming", *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'07)*, New Paltz, New York, October 2007.

-
- [58] Jingdong Chen, Benesty J., Yiteng Huang, "Robust time delay estimation exploiting redundancy among multiple microphones", *Speech and Audio Processing, IEEE Transactions on*, Vol. 11, Issue: 6, pages: 549- 557, Nov. 2003.
- [59] Benesty J., Yiteng Huang, Jingdong Chen, "Time Delay Estimation via Minimum Entropy", *Signal Processing Letters, IEEE*, Vol. 14 , Issue: 3, Pages: 157 – 160, March 2007.
- [60] Benesty J., Jingdong Chen, Yiteng Huang, "Time-delay estimation via linear interpolation and cross correlation", *Speech and Audio Processing, IEEE Transactions on* Vol. 12, Issue: 5, pages: 509- 519, Sept. 2004.
- [61] Talantzis F., Constantinides A.G., Polymenakos L.C., "Estimation of direction of arrival using information theory", *Signal Processing Letters, IEEE*, Vol. 12, Issue: 8, pages: 561- 564, Aug. 2005.
- [62] Talantzis F., Ward D.B., Naylor P.A., "Performance analysis of dynamic acoustic source separation in reverberant rooms", *Audio, Speech and Language Processing, IEEE Transactions on Speech and Audio Processing, IEEE Transactions on*, Vol. 14 , Issue: 4 , pages: 1378 – 139, July 2006.
- [63] Yegnanarayana B., Prasanna S.R.M., Duraiswami R., Zotkin D., "Processing of reverberant speech for time-delay estimation", *Speech and Audio Processing, IEEE Transactions on* Vol.13 , Issue: 6, Nov. 2005.
- [64] Silverman H.F., Sachar J.M., "The time-delay graph and the delayogram - new visualizations for time delay", *Signal Processing Letters, IEEE*, Vol. 12, Issue: 4, pages: 301- 304, April 2005.
- [65] Parisi R., Cirillo A., Panella M., Uncini A., "Source Localization in Reverberant Environments by Consistent Peak Selection", *Acoustics, Speech and Signal Processing*,

ICASSP 2007, IEEE International Conference on, pages: I-37 -40 ,Honolulu, HI, April 2007.

[66] Gustafsson T., Rao B.D., Trivedi M., "Source localization in reverberant environments: modeling and statistical analysis", Speech and Audio Processing, IEEE Transactions on, Vol. 11, Issue: 6, pages: 791- 803 Nov. 2003.

[67] Parisi R., Gazzetta R., Di Claudio E.D., "Prefiltering approaches for time delay estimation in reverberant environments", Acoustics, Speech, and Signal Processing, Proceedings. (ICASSP '02). IEEE International Conference on , May 2002.

[68] Di Claudio E.D., Parisi R., Orlandi G., "A clustering approach to multi-source localization in reverberant rooms", Sensor Array and Multichannel Signal Processing Workshop. Proceedings of the 2000 IEEE, pages: 198 – 201, March 2000.

[69] Di Claudio E.D., Parisi R., Orlandi G., "Multi-source localization in reverberant environments by ROOT-MUSIC and clustering", Acoustics, Speech, and Signal Processing, ICASSP '00, Proceedings 2000 IEEE International Conference on , June 2000.

[70] Adcock F., British Patent, No.130490, 1919.

[71] Guy J.R.F, Davies D.E.N., "Studies of Adcock direction finder in terms of phase mode excitation around circular arrays", Radio and Electronic Engineer, Vol.53, No.1, pp 33-38, January 1983.

[72] Baghdady E.J., "New developments in direction-of-arrival measurement based on Adcock antenna clusters", Aerospace and Electronics Conference, NAECON 1989, Proceedings of the IEEE 1989 National, Dayton, OH, May 1989.

[73] Chan Y.T., Yuan Q., Inkol, R., "A frequency domain implementation of Butler Matrix direction finder ", IEEE, 1999.

[74] Elko G.W., Anh-Tho Nguyen Pong, “A steerable and variable first-order differential microphone array”, Acoustics, Speech, and Signal Processing, ICASSP-97, 1997 IEEE International Conference on, Vol.1, Pages:223–226 , 1997.

[75] <http://en.wikipedia.org/wiki/Steradian>.

[76] Green R. M., Smart W. M., Textbook on Spherical Astronomy, 6 ed. Cambridge, England: Cambridge University Press, 1985.

[77] Abramowitz M., Stegun I. A., Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9 Ed. New York: Dover, 1972.

[78] Widrow B., Stearns D., Adaptive Signal Processing, New Jersey: Prentice-Hall, 1985.

[79] Mendel J. M., Lessons in Digital Estimation Theory. Englewood Cliffs, NJ: Prentice Hall, Inc., 1987.

[80] Robert M. Gray, Lee D. Davisson, “An Introduction to Statistical Signal Processing”, Cambridge University Press, 2004.

[81] Vyacheslav P. Tuzlukov, Signal Processing Noise, CRC Press, 2002.

[82] Prabhakar S. Naidu, Sensor array signal processing. CRC Press, Boca Raton, Fla., 2001.

[83] Cromwell P. R., Polyhedra, New York: Cambridge university press, 1997.

[84] Ballou G. M., Handbook for Sound Engineers, 3rd ed. Boston: Focal Press, 2002.

[85] Hassall J.R., Zaveri K., “Acoustic Noise Measurement, Application of Bruel & Kjaer Equipment”, Bruel & Kjaer, July 1978.

[86] Kenneth E. Barner, Gozalo R. Arce, Nonlinear Signal and Image Processing, CRC Press, 2004.

-
- [87] Trevor J. Cox, Peter D'Antonio, "Acoustic Absorbers and Diffusers: Theory, Design, and Application", Taylor & Francis, 2004.
- [88] Athanasius Kircher, "Musurgia universalis sive ars magna", Corbelletti, Rome, 2 Vols, 1650.
- [89] Kazuhiro Takashima, Hiroshi Nakagawa, Natsu Tanaka, Daiki Sekito "Impulse response measurement system and its recent applications" The Journal of the Acoustical Society of America, , Vol. 120, Issue 5, November 2006.
- [90] Loren Alldrin,"Sound Track: Sound Bounces", Videomaker Enewsletter, November1997, <http://www.videomaker.com/article/3077/>.
- [91] Joong Hyup Ko, Wan Joo Kim, Myung Jin Chung, "A Method of Acoustic Landmark Extraction for Mobile Robot Navigation", IEEE Transactions on Robotics and Automation, Vol.12, No.3, June 1996.
- [92] Aarabi, P., "The Fusion of Distributed Microphone Arrays for Sound Localization", EURASIP Journal of Applied Signal Processing (Special Issue on Sensor Networks), Vol. 2003, No. 4, pp. 338-347, March 2003.
- [93] Richard J. Kozick, Brian M. Sadler, "Distributed source localization with multiple sensor arrays and frequency selective spatial coherence", Statistical Signal and Array Processing, 2000. Proceedings of the Tenth IEEE Workshop on, Pages: 419–423, 2000.
- [94] www.peda.com/polypro.
- [95] Electret condenser microphone cartridge, ceramic microphone receiver cartridge, dynamic microphone cartridge publication Secaucus, NJ: Panasonic Electronic Components Division, Panasonic Industrial Co., vol. 3, 1994.
- [96] www.panasonic.com/industrial/components/PDF.

-
- [97] National Instruments PCI-DAQ Catalogue, NI, Vol. 11, 2002.
- [98] Hoth D.F., "Room Noise Spectra at Subscribers' Telephone Locations", the Journal of the Acoustical Society of America, Volume 12, Issue 3, p. 475, January 1941.
- [99] Silverman H.F., Patterson W.R., Flanagan J.L., Rabinkin D., "A digital processing system for source location and sound capture by large microphone arrays", Acoustics, Speech, and Signal Processing, ICASSP-97., 1997 IEEE International Conference on, Munich, Germany, Apr 1997.
- [100] DiBiase J. H., "A high-accuracy, low-latency technique for talker localization in reverberant environments", Ph.D. dissertation, Providence, RI, Brown University, 2000.
- [101] James W. Cooley and John W. Tukey, "An algorithm for the machine calculation of complex Fourier series", Math. Comput. 19, Page: 297–301, 1965.
- [102] Steven W. Smith, "Digital Signal Processing: A Practical Guide for Engineers and Scientists", California Technical Publishing, 2002.
- [103] Balestro F., et al, "A 3-V 0.5 micro m CMOS A/D audio processor for a microphone array", Solid-State Circuits, IEEE Journal of, Volume: 32 Issue: 7, Pages: 1122 –1126, July 1997.
- [104] Grech I., Micallef J., Vladimirova T., "Low voltage SC TDM correlator for the extraction of time delay", Electronics, Circuits and Systems, 2000. ICECS2000. The 7th IEEE International Conference on, Vol. 1, Pages: 112 –115, 2000.
- [105] Chiang-Jung Pu, Harris J.G.,T "A continuous-time analog circuit for computing time delays between signals", Circuits and Systems, 1996. ISCAS '96, Connecting the World, 1996 IEEE International Symposium on, Vol. 3, Pages: 357-360, 1996.
- [106] Nguyen D., Aarabi, P., Sheikholeslami A. , "Real-time sound localization using field-programmable gate arrays", Acoustics, Speech, and Signal Processing, 2003.

Proceedings (ICASSP '03). 2003 IEEE International Conference on, pages: II - 573-6 vol.2, April 2003.

[107] Halupka D., Mathai N.J., Aarabi P., Sheikholeslami A., "Robust sound localization in 0.18 micro m CMOS", Signal Processing, IEEE Transactions on Acoustics, Speech, and Signal Processing, IEEE Transactions on, Vol. 53, Issue: 6, pages: 2243-2250, June 2005.

[108] Halupka D., Rabi S.A., Aarabi P., Sheikholeslami A., "Real-time dual-microphone speech enhancement using field programmable gate arrays", Acoustics, Speech, and Signal Processing, 2005. Proceedings (ICASSP '05). IEEE International Conference on, pages 149-152, March 2005.

[109] Halupka D., Rabi S.A., Aarabi P., Sheikholeslami A., "Low-Power Dual-Microphone Speech Enhancement Using Field Programmable Gate Arrays", Signal Processing, IEEE Transactions on, Acoustics, Speech, and Signal Processing, Vol. 55, Issue: 7, Part 1, pages: 3526 – 3535, July 2007.

[110] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "'Eye Array" Sound Source Localization" presented at Signals, Systems and Computers, 2006. IEEE Conference Proceedings of the Fortieth Asilomar Conference on, Monterey, CA, October 2006.

[111] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "A Novel Hemispherical Array Sound Source Localization" presented at Signal Processing, 2006. Proceedings of the ICSP '06', IEEE 8th International Conference on, Guilin, China, November 2006.

[112] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "The Audio Surveillance Eye" Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06), Sydney, Australia, November 2006.

-
- [113] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "Eye Array Placement in Enclosed Areas" Proceedings of the IEEE, 20Pth Canadian Conference on Electrical and Computer Engineering (CCECE'07), Vancouver, BC, April 2007.
- [114] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "Difference Closeness Function for Eye Array", Proceedings of the IEEE Pacific Rim Conference on Communications, Computers & Signal Processing (PACRIM'07), Victoria, BC, Aug. 2007
- [115] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "Correlative Closeness Function for Eye Array", to appear in: 2008 IEEE International Instrumentation and Measurement Technology Conference-I2MTC, Victoria, BC, May 2008.
- [116] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "Acoustic source localization with eye array", The Journal of the Acoustical Society of America, Volume 120, Issue 5, November 2006.
- [117] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "Eye Array Dereverberation by Corner Placement", The Journal of the Acoustical Society of America, Volume 121, Issue 5, May 2007.
- [118] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "Alternative Closeness Functions for Eye Microphone Array", The Journal of the Acoustical Society of America, Volume 122, Issue 5, November 2007.
- [119] Hedayat Alghassi, Shahram Tafazoli, Peter Lawrence, "Eye Array: A New Sound Source Localization Method", to be submitted to: IEEE Transactions on Acoustics, Speech and Signal Processing.
- [120] Yiteng Huang, Jacob Benesty, Audio Signal Processing for Next Generation Multimedia Communication Systems, Boston, Kluwer Academic Publishers, 2004.
- [119] www.cs.colorado.edu/~lindsay/creation/eye_stages.html.

Appendix A

Digital Formulation of the MCF Algorithm

Suppose that the total number of microphones on the shell is M and we need at least $N+1$ samples from each signal to achieve a statistically reliable estimation. Therefore at each time step j , we need to employ N previous samples in addition to the sample at the time j :

$$S_i^j = [S_i(j-N) \quad S_i(j-N+1) \quad \dots \quad S_i(j)]^T \quad (\text{A.1})$$

$$\forall i=1,2,\dots,M$$

Since $T_s = \frac{1}{f_s}$, the time steps it takes for sound to travel the hemisphere radius (r) is:

$$n = \frac{T_r}{T_s} = f_s \frac{r}{c} \quad (\text{A.2})$$

where c is the velocity of sound in air, f_s is the sampling frequency and T_r is the time it takes sound to travel r . Thus the approximation equation (3.8) can be written as:

$$\frac{1}{n} [S_0^j - S_0^{j-n}] \cong \frac{1}{n_i} [S_0^j - S_i^j] + N_i^j \quad (\text{A.3})$$

Here, n_i is the time step of traveling sound from m_0 to m_i , and N_i^j is the overall noise and reverberation term.

Obviously we are interested in over-determined situations, on which $N > M$. Since we cannot estimate noise and reverberation term N_i^j , it is neglected here that is the rational for small and moderate amounts of the noise. By increasing N relative to M , we can offset the adverse effect of the noise term reasonably. In other words, an extra measurement filters the data from noise. It also makes our estimation closer to the asymptotic result and boosts the

estimation's unbiasedness. Contrary to frequency domain based SSL methods, increasing the measurement frame (N) can be easily realized in our time domain method without the need for a significant increase of computational power.

Similar, to (3.11) the solution of the standard LMS equation (A.3) is:

$$\frac{n_i}{n} = \{[S_0^j - S_0^{j-n}]^T \cdot [S_0^j - S_0^{j-n}]\}^{-1} \cdot \{[S_0^j - S_0^{j-n}]^T \cdot [S_0^j - S_i^j]\} \quad (\text{A.4})$$

As well, the lens closeness function (3.22) can be written as:

$$\frac{n_i}{n_{i^\perp}} = \frac{[S_0^j - S_0^{j-n}]^T \cdot [S_0^j - S_i^j]}{[S_0^j - S_0^{j-n}]^T \cdot [S_0^j - S_{i^\perp}^j]} \quad (\text{A.5})$$

Note that nominator and denominator in (3.31) are quite similar. In both, the left hand term is the time derivative of the central microphone signal and the right hand term is spatial difference calculated over the i^{th} direction (nominator) and its diagonal $i^\perp$ (denominator). Notice that $i^\perp$ is not only the diagonal of i but also a main direction itself. Therefore we simply compute:

$$F_i^N(j) = [S_0^j - S_0^{j-n}]^T \cdot [S_0^j - S_i^j] \quad \forall i = 1, 2, \dots, M \quad (\text{A.6})$$

once for all directions. The computational load of (A.6) is studied in Subsection 7.3.6 and compared with a pair of TDOA.

Later by dividing each $F_i^N(j)$ by its diagonal counterpart $F_{i^\perp}^N(j)$, we achieve the M desired closeness functions.

Figure 7-3 in Subsection 7.2.1 shows the frontal wireframe view of array with the assigned node numberings.

Moreover, Table 7-1 in Subsection 7.2.1 shows the orthogonal direction pairs for every direction in addition to azimuth and elevation angle of each direction in our test bed geometry.

At a final step we took the average energy of closeness functions on all possible orthogonal pairs for each direction:

$$\overline{F_i^N}(j) = \sqrt{\frac{1}{p_i} \sum_{i^\perp} F_i^N(j)^2} \quad (\text{A.7})$$

where p_i is the number of orthogonal nodes of the node i (Table 7-1).

Later we sort all averaged closeness function estimates ($\overline{F_i^N}(j)$) and choose the first k closeness functions with highest values among them. Theoretically, the minimum suitable number for k is 3, by which we have a spherical triangle. But in practice, we encountered abrupt changes when the sound source direction passes the border of one spherical triangle to the other. This can be avoided by incorporating more cells into our final estimation process (increasing k). Practically, in the case of $k=5$ we achieved a smooth as well accurate result. Thus, in our final algorithm, in addition to the closest inclusive spherical triangle, we consider two additional nodes with higher closeness function values. These two nodes are always neighbours to the spherical triangle with the highest closeness function values.

Having the k directions with highest averaged closeness function in hand, the last step is a simple weighted averaging on the k corresponding node azimuth and elevation angles to calculate the estimated azimuth and elevation angle of sound source direction:

$$\hat{\varphi}_s(j) = \frac{\sum_{i^*} \overline{F_i^N}(j) \cdot \varphi_i}{\sum_{i^*} \overline{F_i^N}(j)} \quad (\text{A.8})$$

$$\hat{\theta}_s(j) = \frac{\sum_{i^*} \overline{F_i^N}(j) \cdot \theta_i}{\sum_{i^*} \overline{F_i^N}(j)} \quad (\text{A.9})$$

where i^* denotes k selected i directions with maximum closeness function values .In addition φ_i and θ_i are fixed topology dependent azimuth and elevation angles acquired from Table 7-1.

The pair $[\hat{\varphi}_s(j), \hat{\theta}_s(j)]$ denotes the final estimated azimuth and elevation angles of the sound source at the current time instance j .

MCF Algorithm Flowchart

The flowchart of our sound source localization method utilizing multiplicative closeness function described by equations (A.5) or (3.22) is shown in Figure A.1:

Figure A-1 Flowchart of the MCF based eye array processing

Appendix B

Far Field Assumption

In this research, the sound source was assumed to be a point source and the sound wave was assumed to be planar. Practically the sound source is not a point but can be considered to be a spherical source radiating spherical waves which are decreasing in power proportional with inverse of the square of the distance (Figure B-1).

Figure B-1 Spherical radiation of spherical sound source

For an array with limited dimensions(R), placed far away from this source(L) the wavefront looks planar and the variation of sound level is negligible.

In array theory the common rule of thumb (Rayleigh criterion) to determine the distance at which the far field approximations begin to be valid is [1, 2]:

$$L \geq \frac{2R^2}{\lambda} \quad (\text{B.1})$$

Here, R corresponds to the array spacing and λ is the wavelength.

Assuming a hemisphere with the radius $R=0.34$ meters and signal bandwidth of up to 8 kHz, the minimum L would be 1.4 meters.

This distance will typically be exceeded in applications such as conference rooms, camera pointing devices or speaker localizations. The smaller the array radius, the smaller the minimum required distance would be.

Appendix C

Eye Analogy

There are some similarities between our localization array and the eye. The most obvious one is the main localization strategy which in both cases is based on location estimate of some activated sensors distributed on a hemispherical surface. Each light ray enters the hemispherical eye chamber through either a pinhole or a lens and finally has an effect on the corresponding retina sensor in its direction of arrival. In the eye structure, the direction of arrival of the light ray is detected by the location of the activated retinal sensors. Likewise, the sound source direction of arrival is carried out by the direction of the maximum outputs of the microphone cells. Figure C-1 shows two different types of eye mechanisms that exist in living creatures.

In a pinhole eye, light passes through a single hole in the center and affects the retinal sensor located on its path. In the eye microphone array with the pinhole closeness function, sound passes the central microphone to the shell microphone that is located in the direction of the sound, and builds up a high closeness function. As stated earlier in this chapter, with the pinhole closeness function, due to small aperture, the resolvability is low.

Adding a lens to a pinhole eye increases the resolving ability of the eye by increasing the effective aperture of measurement and creating a focal point. A comparable situation occurs in the eye microphone array. By adding the third orthogonal microphone to the two-microphone (pinhole) cell we have increased the aperture of measurement and created a lens-like focusing system, which increases the resolution.

Finally, the retinal photoreceptors in the eye have been arrayed in triangular-tessellated hexagonal meshes. Likewise in our topology we have triangular-tessellated pentagon meshes. These analogies persuade us to call our localization array and strategy, an “eye array sound source localization”.

Figure C-1 Eye mechanism [119]