Acoustic Imaging Using a 64-Node Microphone Array and Beamformer System

by

Feng Su

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Applied Science

in

Human-Computer Interaction

Carleton University
Ottawa, Ontario

© 2015
Feng Su

Abstract

Acoustic imaging is difficult to achieve in environments with a large amount of noise and reverberation. Microphone arrays, as a branch of array signal processing, offer an effective approach to obtaining a clean recording of desired acoustic signals in these environments. In this thesis, we have designed, implemented, and evaluated a 64-node microphone array system for acoustic imaging. We have applied a delay-and-sum beamforming algorithm for sound source amplification in a noisy environment, and have explored the uses of the array and beamformer by generating the sound intensity map to reconstruct the acoustic scene of interest. Our experimental results show a mean error of 1.1 degrees for sound source localization, and a mean error of 13.1 degrees for source separation. In addition, we also used the system to image seven different materials with audible sound, and obtained their reconstructed acoustic maps as well as frequency response curves, from which we are able to detect the differences between textures based on their acoustic response powers.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Chris Joslin, for the continuous support of my Master's study, and his guidance throughout this thesis. My sincere thanks also goes to my fellow labmates, especially Rufino R. Ansara, for his assistance in data collection. I would also like to thank my defense committee, for their insightful comments and suggestions. Finally, thanks to my family: my parents and grandparents who have encouraged and supported me throughout my study and life.

Table of Contents

Abstract ..... **ii**

Acknowledgements ..... **iii**

Table of Contents ..... **iv**

List of Figures ..... **vii**

List of Acronyms ..... **ix**

1 Chapter: Introduction **1**

1.1 Applications of Acoustic Imaging 3

1.2 Research Questions 6

1.3 Thesis Overview 8

1.4 Contributions 10

2 Chapter: Related Works **13**

2.1 Microphone Array Technology 13

2.1.1 Speech Recognition 17

2.1.2 Sound Source Localization and Separation 18

2.1.3 Range Detection and Navigation System 21

2.1.4 In-air Gesture Sensing 22

2.1.5 Object Detection 23

2.2 Acoustic Imaging Strategy 24

2.2.1 Lens Based Strategy 24

2.2.2 Time-Reversal Strategy 26

2.3 Array Signal Processing 26

2.3.1 Static Beamforming 27

2.3.2 Adaptive Beamforming 28

3 Chapter: Methodology 30
3.1 Speed of Sound 32
3.2 Far-field and Near-field Cases 32
3.3 Problem Description 33
3.4 Delay-and-Sum Beamforming 34
3.5 Signal-to-Noise Ratio 36
3.6 Beam Pattern 37
3.7 Sensor Spacing 40

4 Chapter: System Design 43
4.1 Hardware Design 44
4.1.1 The Sensors 44
4.1.2 Array Geometry 45
4.1.3 Data Acquisition System 48
4.2 Software Design 49
4.3 System Implementation 52
4.4 Cost 54

5 Chapter: Experiments and Results 55
5.1 Testing Environment 55
5.2 Array Beam Pattern 56
5.3 Beamforming 57
5.4 Source Localization and Separation 59
5.4.1 Sound Source Localization 59
5.4.2 Sound Source Separation 62
5.5 Imaging of Different Materials 63

6 Chapter: Discussion 69
6.1 Performance of Source Localization and Separation 69

6.1.1 Localization Performance 69
6.1.2 Separation Performance 71
6.1.3 Image Resolution 72
6.2 Acoustic Response of Different Materials 73
6.3 Summary 75
6.4 Potential Application for Mobile Robot Navigation 77
7 Chapter: Conclusions 79
7.1 Summary of Findings 79
7.2 Limitations and Future Work 80
References 82

List of Figures

Figure 1 Problems with microphone array signal processing. 6

Figure 2 A photograph of the LOUD 1020-node microphone array. 15

Figure 3 Photograph of a typical sonic crystal sample taken from underneath [48]. 25

Figure 4 Illustration of the delay-and-sum beamforming process. 31

Figure 5 Illustration of an equispaced linear array, where the source *s(k)* is located in the far field, the incident angle is θ, and the spacing between two neighboring sensors is d. 38

Figure 6 Beam pattern of a ten-sensor array when $\theta = 90°$, $d = 8$ cm, and $f = 2$ kHz: in Cartesian coordinates (left) and in polar coordinates (right). 40

Figure 7 Beam pattern (in polar coordinates) of a ten-sensor array when $\theta = 90°$, $d = 24$ cm, and $f = 2$ kHz. 41

Figure 8 The schematic diagram of the 64-node microphone array hardware design. ... 43

Figure 9 The microphone module selected for our array (left) and its frequency response curve (right). 44

Figure 10 Experimental results from the LOUD microphone array [31, 76]: peak SNRs for one representative recording (left), and experimental recognition accuracies (right). 47

Figure 11 Functional block diagram of the PMC66-16AI64SSA data acquisition board. 48

Figure 12 Block diagram of algorithms for our acoustic imaging microphone-array system. 50

Figure 13 Azimuth and elevation angle with respect to the uniform rectangular array. 51

Figure 14 A photograph of the 64-node microphone array system. 53

Figure 15 Testing environment for the microphone array beamforming system. 55

Figure 16 Beam patterns of rectangular arrays with different geometries: (a) 1 by 2, (b) 2 by 2, (c) 3 by 3, (d) 4 by 4, (e) 5 by 5, (f) 6 by 6, (g) 7 by 7, (h) 8 by 8..... 56

Figure 17 Array gains as the number of microphones is increased. The look direction is at 0 Az and 0 El..... 57

Figure 18 Comparison of the collected signal before and after beamforming (left), and their corresponding spectrums (right). 58

Figure 19 Setups for sound localization at 64 positions (top) and the corresponding results (bottom). 60

Figure 20 Separation of two source signals whose spacing ranges from 6.3 cm to 36.3 cm..... 62

Figure 21 Setup for the imaging experiment. The speaker was located 5 cm to the left edge of the array, and the object was 30 cm in front of the array..... 64

Figure 22 Reconstructed acoustic images of 7 different materials using a range of frequencies. 67

Figure 23 Frequency response curves of 7 different materials. 68

Figure 24 Beam patterns with the increase of the frequency from 1 kHz to 7 kHz. The look direction of the array is at 0 Az, 0 El. 72

List of Acronyms

ADC	Analog-to-Digital Converter
ASR	Automated Speech Recognition
Az	Azimuth
BSS	Blind Source Separation
CAD	Computer-Aided Drafting
CSAIL	Computer Science and Artificial Intelligence Laboratory
DOA	Direction of Arrival
DSBF	Delay-and-Sum Beamforming
DSP	Digital Signal Processor
El	Elevation
FFT	Fast Fourier Transform
FPGA	Field-Programmable Gate Array
GSC	Generalized Sidelobe Canceller
HCI	Human-Computer Interaction
HMA	Huge Microphone Array
ICA	Independent Component Analysis
ISA	Industry Standard Architecture
LEMS	The Laboratory for Engineering Man/Machine Systems
LOUD	Large acOUstic Data Array
MDR	Mini D Ribbon
MEMS	Micro-Electro-Mechanical Systems

MUSIC	MUltiple SIgnal Classification
MVDR	Minimum Variance Distortionless Response
NBSFC	Nullspace-Based Sound Field Control
PCB	Printed Circuit Board
PCI	Peripheral Component Interconnect
PMC	PCI Mezzanine Card
PVC	Polyvinyl Chloride
RADAR	Radio Detection and Ranging
SLAAM	Scalable Large Aperture Array of Microphones
SNR	Signal-to-Noise Ratio
SONAR	Sound Navigation and Ranging
SPL	Sound Pressure Level
TDOA	Time Difference of Arrival
TOF	Time-of-Flight
URA	Uniform Rectangular Array
US	Ultrasound
WER	Word Error Rate

1 Chapter: Introduction

Acoustic imaging can be described as a method for recording and reconstructing the amplitude distribution of a propagating sound field in a given plane. It is a field which has grown considerably over the past decade, and has been widely developed for important applications such as medical ultrasonography, non-destructive evaluation, and underwater sonar. While acoustic waves have been shown to be effective for imaging underwater and in the body, their use in-air is difficult since the propagation speed of sound in air is relatively low and interference from noise and echo is high. So far, a large amount of work has been done to explore possible solutions for in-air acoustic imaging.

Since the 1960's, early research has been mainly focused on obtaining acoustic images in a way more similar to optics, which later has been extended to include holographic techniques. This has led to the establishment of a new discipline - acoustic holography [1]. In the field of optics, holography has become identified with three-dimensional reconstruction. However, this is not the case with acoustic holography. The problem here is that acoustic holograms are recorded at the wavelength of sound, but are then reconstructed at the wavelength of visible light. This scaling down in wavelength will consequently introduce a large distortion that makes the exact three-dimensional reconstruction impossible. Although various methods to reduce this distortion have been proposed and there are also other techniques that have no direct similarity to optical

holography, still due to their complex set-ups, acoustic holography will not be the most practical way to generate acoustic image for now.

More recently, new techniques have been developed to record and generate acoustic images in air. Microphone arrays are capable of providing spatial information for incoming acoustic waves, as they can capture key information that would be impossible to acquire with single microphones. Acoustic imaging microphone arrays (sometimes referred to as "acoustic cameras" [39]) often contain a camera which is usually located at the center of the array. An acoustic map, generated using the microphone data, is overlaid as a transparency over the camera image. With certain array signal processing techniques it is possible to localize individual sound sources in the recorded sound field and their emitted sound pressure level (SPL) from that acoustic image.

While a wide frequency range is available for acoustic imaging, ultrasound (US) becomes one of the most widely used imaging technologies as it offers a much higher frequency (usually above 20 kHz) and can easily penetrate opaque media. This is why it has been successfully applied for imaging in medical and underwater sonar applications, and also frequently needed for testing of multi-layered objects in industrial production [21]. Medical ultrasonic imaging has been used to image the human body for over half a century. It is fast, portable, free of radiation risk, and relatively inexpensive when compared with other imaging modalities, such as magnetic resonance and computed tomography. Furthermore, ultrasound images are tomographic, i.e., they can offer a "cross-sectional" view of anatomical structures. The images can be obtained in real time,

thus providing instantaneous visual guidance for many interventional procedures. To improve image quality, ultrasound contrast agents such as microbubbles are often used. They have been successfully employed with a wide range of imaging techniques, and become the subject of a broad and rapidly developing field of research [22, 23].

Modern medical ultrasound system is performed primarily using a pulse-echo approach with a brightness-mode (B-mode) display. The basic principles of acoustic imaging are similar to the B-mode ultrasonography, which involves transmitting small pulses of ultrasound waves from a transducer into the body, and the echo signals returned from many sequential coplanar pulses are processed and combined to generate an image [54]. While high-frequency ultrasound equipment (up to 20 MHz) generate images of high resolution, their costs are relatively high for general consumer applications.

Although ultrasound is capable of imaging in human body and underwater, it can be disrupted by air or gas, and therefore is not the most ideal imaging technique for applications that require to be operated in air. This has led us to explore other imaging modalities, such as using audible sound.

1.1 Applications of Acoustic Imaging

Conventionally, acoustic imaging can be divided into active imaging (where a transmitter produces acoustic energy which is either reflected from, or transmitted through, the object of interest) and passive imaging (where the object itself is the source of the acoustic energy) [56]. Its applications range from the submillimeter distances of acoustic

microscopy to hundreds of miles in some passive sonar applications. Correspondingly, the frequencies used vary from gigahertz to a few hertz, and the wavelengths from a few micrometers to thousands of meters.

The most typical application for acoustic imaging arrays is teleconferencing. Alignment of the optical camera with the array is commonly used in conference room applications. The cameras may be used to identify the location of a person's head. This information is then used to steer the microphone array towards the person for speech enhancement. However, such system requires special camera or image processing techniques, and in most cases the location of the person in the room needs to be fixed. So far less attention has been given to the uses of the microphone array itself to extract the locations of individuals through the sound source images.

Mobile platforms that operate in everyday environments not only need to detect obstacles in their surroundings, but should also become aware of the presence of other objects. Detecting objects in the surroundings of a robot is crucial for control of its awareness and navigation. Based on the detection, the orientation and trajectory of the object can be estimated and the system can respond meaningfully, such as stepping out of the way of the object. Compared to optical and radar based systems, acoustic imaging provides a simple and cheap sensor alternative that allows for relatively precise range as well as angular information. Using an acoustic array, objects can be easily detected in the environment and a 3D image of reflections in the surrounding scene can be created [11,

12]. Object detection based on such a system can greatly enhance the overall system reliability.

The importance of acoustic imaging technologies has also been recognized in the human-computer interaction (HCI) community. There is a rapidly growing body of work that explores applications of this emerging image processing technology, including the possibility of creating novel 3D gesture recognition, developing interactive tools for the design and rapid fabrication of interactive systems and many others.

To summarize, we briefly list the typical applications of microphone arrays and acoustic imaging:

- Teleconferencing,
- Hands-free acoustic human-machine interfaces,
- Command-and-control interfaces,
- Speech enhancement and recognition
- Video games,
- Mobile robot navigation systems,
- High-quality audio recordings,
- Acoustic surveillance (security and monitoring),
- Acoustic scene analysis,
- Sensor network technology.

We can see that the number of applications is enormous and can be expected to grow as time goes.

1.2 Research Questions

In this thesis, we address two main research questions: The first one is how to achieve acoustic imaging in air, both cheaply and accurately, with microphone arrays, especially in an environment where noise, echo and reverberation is high. The second question is how to utilize this imaging array system to measure the acoustic responses of different entities, and how the results can be affected by different frequencies and textures.

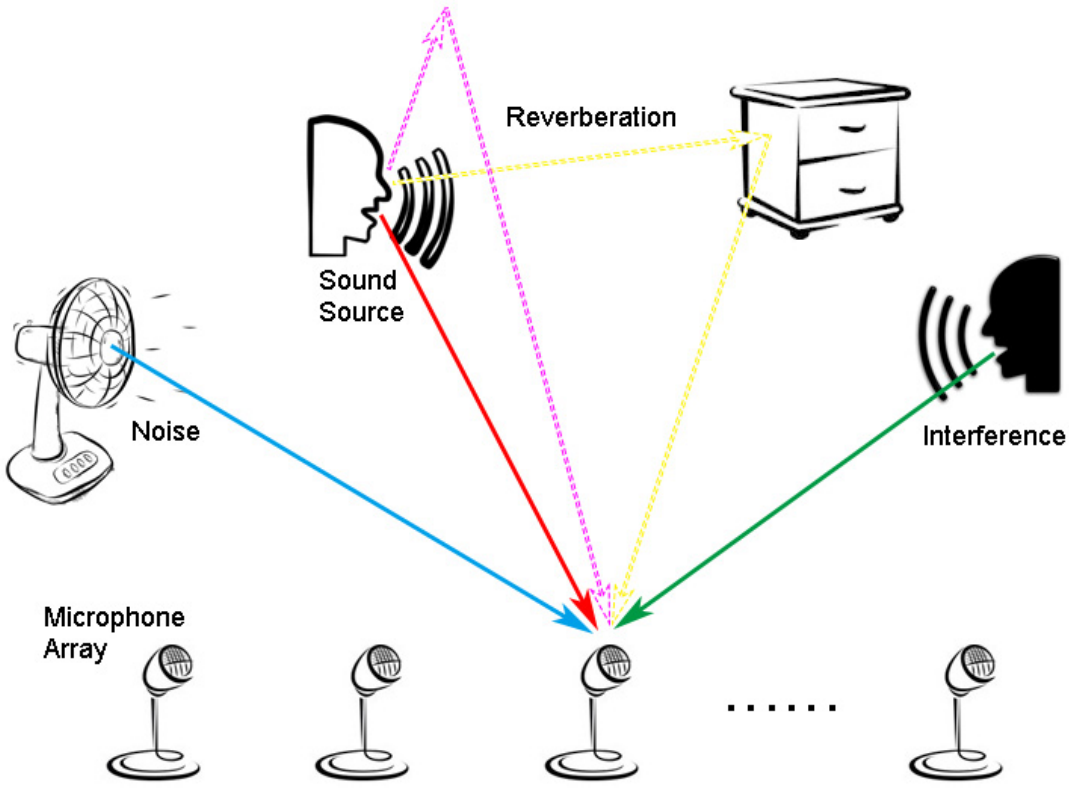

Figure 1 Problems with microphone array signal processing.

To reconstruct the acoustic scene with microphone arrays, we need first to recover the source signals from the observed ones, and this usually includes noise reduction and dereverberation. We illustrate these problems in Figure 1, where all the signals received by the microphones will pass through certain filters that need to be optimized according to one of the above-mentioned problems.

The objective of a noise reduction algorithm is to estimate the desired source signal from its corrupted observations which are due to the effects of an unwanted additive noise. With a microphone array, we should be able to reduce the noise without affecting much the sound source signal.

In a room with furniture and multiple other objects, the signals that are received by microphones from a sound source contain not only the direct-path signals, but also attenuated and delayed replicas of the source signal due to reflections from boundaries and objects in this room. This multipath propagation effect introduces echoes and spectral distortions into the observation signals (termed as reverberation), which may severely deteriorate the source signal causing quality and intelligibility degradation. Therefore, dereverberation is required to improve the quality of the source signal. Great efforts have been made in the past decades to find practical solutions with a microphone array.

Finally, the reflection and scattering of the source signal is influenced by the geometry and materials of the walls and other scatterers in the environment. Although we tend to avoid such signals for source localization and separation, they are of great interest with

acoustic navigation systems since they contain the key information for distinguishing different objects (such as human and non-human) in front of the detecting device.

All of the aforementioned problems are very difficult to solve no matter the size, the geometry, or the number of elements of the array. Therefore, to address the first research question, our objective is to find the proper array signal processing algorithm for in-air acoustic imaging, and explore the design and implementation of a microphone array. For the second research question, our goal is to image typical objects with different textures using audible sound, and find out how their response powers vary with textures. We hypothesize that different textures will have different acoustic responses with the increase of testing frequencies.

1.3 Thesis Overview

This thesis is organized as follows:

Chapter 2 provides a review of microphone array technology. We describe some current microphone array applications which are related to our work. This includes speech recognition, source localization and separation, range detection and navigation, in-air gesture sensing, and object detection, as well as some typical strategies for acoustic imaging. We also discuss some array processing algorithms, which include static and adaptive beamforming.

In Chapter 3 we introduce the algorithms used in this thesis. We begin by outlining some of the basic concepts of beamforming algorithm, and then discuss in further details the delay-and-sum beamformer and also give the simulation results of its beam pattern in order to illustrate its performance. We conclude this chapter by outlining some requirements for array design.

In Chapter 4 we apply the proposed algorithms to a 64-node microphone array for acoustic imaging. We outline the process of designing both the hardware and software for the acoustic imaging array. This includes the components and tools we used, array geometry, and the program developed for the beamformer. We also provide the cost for building such microphone array system.

Chapter 5 focuses on evaluating the performance of the array system and verifying our hypothesis. We simulate the beam pattern of this 64-node array, and test it with sound source localization and separation. We then examines the acoustic responses of different materials using the array and a speaker, and compare their frequency response curves as well as their corresponding acoustic images. In Chapter 6 we discuss these results in more depth and look at its potential for mobile robot navigation.

Finally, Chapter 7 summarizes the major results and contributions of this thesis and highlights some directions for future research.

1.4 Contributions

Acoustic imaging with microphone arrays has been an ongoing topic for more than a decade. Compared to other array systems presented in the current literature, we highlight the following key characteristics of our system that contribute to its capability in acoustic imaging applications:

- High accuracy: Our system is capable of localizing sound source with an average error of 1.1 degrees. It can also separate two sound sources with an average error of 13.1 degrees, which achieves significant improvement over previous small-scale microphone array localization system.
- Low power consumption: All of the microphones are powered by a 3.3V/0.014A DC power supply, which is 46.2 mW in total. The output power of the speaker is 200 mW. Thus, the overall system power consumption is 246.2 mW. Compared to optic-based system such as Kinect (whose power demand is around 12 W), our system consumes approximately 98% less power.
- Large array aperture: Our microphone array has 64 elements (8×8), with a sensor spacing of 2.3 cm (18.4×18.4 cm in total). The measured angular detection range is approximately 55 degree in azimuth, and 50 degree in elevation. In comparison with most of the currently-published works for acoustic imaging, we are among the first to test the performance of such large-scale microphone array.
- Robustness in the presence of noise and reverberation: Our array system can accurately locate and separate sound source signals in a noisy and reverberant environment, which proves its suitability for practical applications (such as teleconferencing) in everyday surroundings.

- Cost effectiveness: The overall cost for the system is around $630, or $1.745/square centimeter installed, which is relatively cheaper, especially compared to commercial products (whose cost-per-unit-area are usually above $5/sq cm).
- Simple, efficient, easy to build and use: Our system is constructed entirely using commodity, off-the-shelf audio modules and 3D printing technology. The programs we developed allow for immediate visualization of the raw data as well as the obtained acoustic images.

Our work provides a thorough design guideline for the implementation of a 64-node microphone array system, which has been successfully applied for sound source localization and separation as well as in-air acoustic imaging with audible sound. Although we do not implement other complex adaptive beamforming algorithms, we show that the delay-and-sum beamforming technique presented is capable of good performance even in highly noisy and reverberant environment.

We also examine the potential of our imaging array system for detecting objects with different textures. We obtained the acoustic images of 6 different materials and a human hand, and compared their frequency responses from 1 kHz to 7 kHz. To our knowledge, few works have ever measured such frequency responses within the range of audible sound, let alone in the context of acoustic imaging. From the results, we observed that while materials with smooth textures have similar frequency response patterns, the responses from rough textures (such as cardboard and human skin) present some significant differences at certain frequencies. Compared to smooth textures, their

response powers are relatively weak from 3.5 kHz to 6.5 kHz. This demonstrates our system's effectiveness of detecting the differences between textures. We show that audible sound can be a cheap, low-power alternative technology for acoustic imaging.

2 Chapter: Related Works

In this chapter, we survey the state of the art in three relevant fields: microphone array technology and its applications, acoustic imaging strategy, and array signal processing.

2.1 Microphone Array Technology

A microphone array is a composition of spatially distributed microphones. Microphone arrays feature the capability of obtaining the actual three-dimensional position of sound sources by estimating several direction-of-arrivals given geometrical considerations [55]. In an acoustic enclosure, such as an auditorium, conference room, concert hall, or automobile, ambient noise and reverberation degrade source signals. Microphone arrays can be steered in software toward a desired sound source, filtering out undesired sources. When an appropriate level of computational power is available, microphone arrays can also track a desired source around a space as the source moves.

Over the past two decades, microphone arrays have been increasingly used for sound source separation and amplification, and since late 1980s, they have been applied for capturing audio in difficult acoustic environments. Flanagan et al. [70] at AT&T Bell Lab experimented with large microphone array designs in the late 1980's. They designed and tested a rectangular array consisting of 63 microphone elements arranged into 9 columns and 7 rows on a 1 square meter panel. A later iteration of the system included 400

microphones [71]. This system was deployed in an auditorium for directional audio capturing to support remote conferencing in the reverberation and noisy conditions.

The Huge Microphone Array (HMA) project from the Laboratory for Engineering Man/Machine Systems (LEMS) group at Brown University designed an array consisting of 512 microphones [69], and was deployed in a lab space of 690 square feet. HMA followed a component design approach. The 512 microphone nodes were grouped into 32 distinct modules. Each microphone module consisted of a printed circuit board (PCB) with 16 microphones, an independent analog-to-digital converter (ADC) and a dedicated digital signal processor (DSP). The DSP was responsible for single channel processing such as frequency transformation and bi-channel processing such as delay computations. The 32 modules were connected to a central processing unit via optical fiber cables. Custom DSPs and a load-and-go operating system were designed and built for the array to accommodate an estimated 6 GFlops of computation rate. HMA was used for a number of array processing tasks, including robust acoustic beamforming, and single and multi-source localization.

The Large acOUstic Data Array (LOUD) project by MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) built an array with 1020 microphones and holds the world record for the most number of microphones on a single array [31, 76]. The LOUD was an rectangular array and the microphones were arranged uniformly with a spacing of 3 cm on a panel of approximately 180 cm wide and 50 cm high (Figure 2). The 1020 nodes were attached to 510 PCBs. Each PCB module contained two

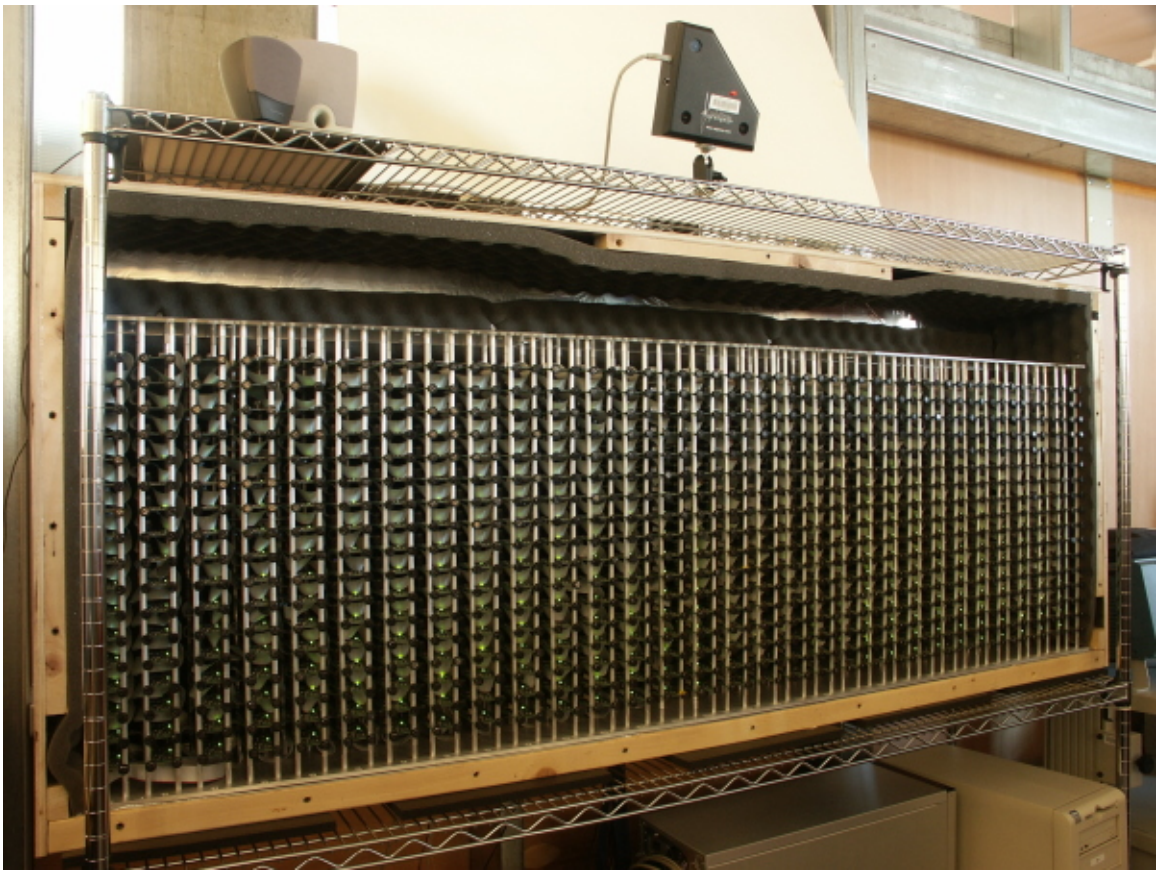

Figure 2 A photograph of the LOUD 1020-node microphone array.

microphones, a stereo ADC and a small cooling component. The ADC sampled analog inputs at 16 kHz, generating 24-bit serial data. The PCB modules were assembled in a LEGO-like fashion: a chain of 16 PCB modules feed into a single input on a connector board using time-division multiplexing, 4 connector boards were used and each hosted 8 such PCB chains. The array produced a total data rate of 393 Mbits/sec. To accommodate the high bandwidth, a customized parallel processor, based on the Raw industry standard architecture (ISA) [75], was designed and used. LOUD was primarily evaluated on automated speech recognition (ASR) tasks. It demonstrated significant word error rate

(WER) improvement over single far-field microphones for speech recognition in both normal (89.6% WER drop) and noisy conditions (87.2% WER drop).

As microphone array technologies have become cheaper and increasingly accessible, there has been growing interest to use such setups for capturing contextualized audio events for building context-aware applications. More recently, rapid advances in processor technologies have propelled small-scale microphone arrays into many consumer electronic products such as smartphones and personal gaming devices. Most smartphones nowadays have at least two microphones for noise cancellation. Some emerging phones, such as Lumia 925 [28], even have three microphones, which will enable more microphone array based applications such as automatic voice tracking.

Commercial products for audio beamforming have also been developed by various companies. For example, the Microsoft Kinect [27] sensor contains a linear microphone array which was originally designed for beamforming, a technique used to amplify sound from one direction and suppressing sound coming from other directions. The microphone array features four microphone capsules with each channel processing 16-bit audio at a sampling rate of 16 kHz. It enables the device to conduct ambient noise suppression so that the player can interact with the game via voice recognition. Other companies, such as Acoustic Camera [29], developed a PC-based beamforming system that utilizes sound acquisition arrays ranging from few tens to more than a hundred elements. Polycom and Microsoft presented the CX5000 [30] unified conference station that leverage microphone array technology in audio conferencing applications [72]. In a recent

research project, Theodoropoulos et al. [33] implemented a custom architecture as a multicore reconfigurable processor for audio beamforming systems, their proposed solution can extract up to 16 audio sources in real time under a 16-microphone setup.

There is currently a significant amount of research and numerous applications which use microphone array for communications, detection and analysis. We list some of the research topics that are related to our work below.

2.1.1 Speech Recognition

Communication through speech has been extensively explored as a method for making human-computer interaction more natural. However, speech recognition performance degrades significantly in distant-talking environments, where the speech signals can be severely distorted by additive noise and reverberation. In such environments, the use of microphone arrays has been proposed as a means of improving the quality of captured speech signals.

More recently, the demand for hands-free speech communication and recognition has increased and as a result, newer techniques have been developed to address the specific issues involved in the enhancement of speech signals captured by a microphone array [63]. As mentioned before, one of the most famous implementations is the LOUD array, which was part of the MIT Oxygen project [32]. The LOUD system is based on the delay-and-sum beamforming algorithm for sound source amplification in a noisy

environment. Experimental results suggest that utilizing such a large microphone array can dramatically improve the source recognition accuracy up to 90.6%.

2.1.2 Sound Source Localization and Separation

Another major functionality of microphone array signal processing is the estimation of the location from which a source signal originates. In acoustic environments, the source location information plays an important role for applications such as automatic camera tracking for videoconferencing and beamformer steering for suppressing noise and reverberation. Estimation of the source location, which is often called source-localization problem, has been of considerable interest for decades. It is accomplished by utilizing differences in the sound signals received at different observation points to estimate the direction and eventually the actual location of the sound source. Two or three dimensional microphone arrays are required to estimate the angle of arrival or the position in Cartesian coordinates of the source. For the two related problems of estimating the number of sources and localizing multiple sources, several interesting algorithms, such as MUltiple SIgnal Classification (MUSIC) [3], exist for narrowband signals [2]. However, researchers have just started to investigate these problems for broadband sources.

Source localization is a fundamental task for microphone array systems and extensive research on localization strategies exists. Zhang et al. [38] provided simulation results of a signal source localization system based on an 8-element uniform linear microphone array and MUSIC algorithm. Pei et al. [6] presented an approach for locating a sound

source using the small-scale linear microphone array on Kinect. Their positioning results showed an average error of 0.25 meters along the horizontal axis and 0.53 meters error along the vertical axis. Goseki et al. [4, 5] proposed a method of visualizing sound pressure distribution by combining microphone array processing with camera image processing. Meyer et al. [37] further described an approach to visualize sound distribution on 3D models via microphone arrays and a laser scanner. Huang [41] presented a real-time recursive algorithm based on the classical state observer in linear control theory to identify the main noise source of an aircraft model for wind tunnel tests. Sun [73] examined the use of a scalable microphone array system known as Scalable Large Aperture Array of Microphones (SLAAM) to extract the locations of individuals in real-time. The system was built with 36 microphones that hung at equal distance from the beams and was deployed in an open semi-structured lab covering a physical space of roughly 1000 square feet. It achieved high accuracy, low latency human speech localization and small-group conversation analysis.

In addition to 2D planar arrays, 3D geometries such as spheres have also been used to localize the sound sources with a variety of techniques. O'Donovan et al. [34] showed through a spherical microphone array that the passive localization of sound sources and their reflections in a concert hall is possible. Legg and Bradley [39] presented a calibration technique for an acoustic imaging spherical microphone array, combined with a digital camera. This technique did not require prior knowledge of microphone positions, inter-microphone spacings, or air temperature. It was applied to a spherical array with 72 microphones and three cameras. The calibration results were then applied to acoustic

imaging for source localization using beamforming and CLEAN-SC algorithms. A mean position difference of 6.6 mm of the sound source coordinates in the acoustic maps was obtained compared to the coordinates obtained using optical computer vision techniques.

In source separation with multiple microphones, the problem here is to separate different signals coming simultaneously from different directions. All the approaches are blind in nature since there is not usually access to either the acoustic channels or the source signals. Independent component analysis (ICA) [77] is the most widely used tool for the blind source separation (BSS) problem, since it takes fully advantage of the independence of the source signals. For example, Turqueti et al. [7] provided the first results on the use of a 52 microphone micro-electro-mechanical systems (MEMS) array, embedded in a field-programmable gate array (FPGA) platform as a source separation system utilizing the ICA technique. They further tested their system to monitor and localize the heart sound, and observed two very distinct heart beat patterns on the obtained acoustic image [58]. Similarly, Kajbaf and Ghassemian [64] proposed a new imaging method for heart sound segmentation. Their system was based on a smaller 3 by 3 microphone array using the delay-and-sum beamforming technique.

While most of the algorithms based on ICA work very well when the signals are mixed instantaneously, they do not perform that well in a reverberant environment. Although a significant amount of progress has been made recently, it is still not clear how and to what degree this can be useful in speech and other acoustic applications.

2.1.3 Range Detection and Navigation System

Information on the distance to the target is very important to achieve the practical use of hands-free speech interfaces and nursing-care robots. For instance, Strakowski et al. [9] used sound to develop an obstacle detecting aid for visually impaired people. Their system operated at 18.4 kHz, and used phase beamforming with an array of 64 microphones. Harput and Bozkurt [10] also proposed a mobility aid for the blind based on ultrasound. The system contained six transmitting and four receiving elements, which were organized in linear arrays for imaging in the horizontal plane. Bedri et al. [36] implemented a method to sense stationary diffuse sound reflecting objects using active, in-air sensing, with a pair of microphones and speakers which were moved in a vertical plane. This technique was applied to detect a mannequin hidden around the corner. In addition, Lang et al. [74] used a two-microphone array as part of a multi-modal person tracking system on a mobile robot.

In the field of acoustic imaging, there also exist some works on range detection. Miyake et al. [8] demonstrated an acoustic range imaging system based on the phase interference method using audible sound. They applied the beamforming technology to range spectra calculated from a linear microphone array to estimate both distance and direction of the target in the short-range. In ultrasonic imaging, however, the Doppler Effect, which is caused by the movement of the target, causes a frequency shift on the reflected sound, and usually results in object detection failure. To overcome this, Maeda et al. [24] constructed an experimental system which comprised of 128 MEMS microphones and an

FPGA for Doppler ultrasonic 3D imaging. By utilizing a log-step multicarrier signal as the transmitter wave, they succeeded in obtaining 3D imaging of a moving object.

2.1.4 In-air Gesture Sensing

Gesture is becoming an increasingly popular means of interacting with computers. However, it is still relatively costly to deploy robust gesture recognition sensors in existing mobile platforms. As an alternative, sonic gesture sensing has been shown to be effective for recognizing a variety of in-air gestures for controlling interfaces. Current technologies, however, have focused on separate transducers and receivers that require custom hardware. For example, Kalgaonkar et al. [40] developed a device, which was based on the Doppler Effect, to recognize one-handed gestures in 3D space using low-cost ultrasonic transducers that emit a 40 kHz tone. They placed one transmitter and three receivers in a triangle pattern where gestures could be performed and sensed. In addition, Adib et al. [65, 66] presented a motion tracking system using radio reflections that bounce off a person's body. Their system was based on multiple transmit and receive antennas mounted on a foldable platform and arranged in a single vertical plane. By estimating the time-of-flight (TOF) of received signals, they were able to recognize concurrent gestures performed in 3D space by multiple users.

To utilize the most ubiquitous components in computing systems, Gupta et al [35] demonstrated that using existing speakers and microphones on commodity devices such as laptops and mobile phones, movement and gesture recognition is possible by sensing

the Doppler shift of reflected sound. This solution worked across a wide range of existing hardware to facilitate immediate application development and adoption.

2.1.5 Object Detection

Object detection is commonly based on optical imaging sensors. For example, LIDAR and Kinect use infrared light, while stereo cameras use visible light. These systems require hardware operating at high sampling frequencies, precise calibration, and they dissipate significant power. Object detection by sonic or ultrasonic means is attractive because of its relatively low power consumption, and simpler, low-rate hardware. Additionally, it could complement light-based detection in scenarios where light fails, such as mirrors, windows and glass walls, imaging through thin fabric, or spaces filled with smoke.

Moebus and Zoubir [11] studied ultrasound imaging in air for object detection, and discussed its suitability for biometric applications such as human presence detection [12]. Their system was based on beamforming with a synthetic 2D array of 400 acoustic receivers. Dokmanic and Tashev [13] designed a simple ultrasonic device with eight MEMS microphones and eight piezo transducers operating at 40 kHz, for acquiring images in both azimuth and elevation. To deal with imperfect beamforming, they proposed to combine the beamformer with sound source localization algorithms (MUSIC). They obtained depth images that revealed the pose of a human subject. This suggested that ultrasound could be used for skeletal tracking, or more generally, human-

computer interaction. However, due to its high attenuation nature, the use of ultrasound in air, especially in the presence of noise and reverberation, is still limited.

2.2 Acoustic Imaging Strategy

2.2.1 Lens Based Strategy

Similar to light, sound can be focused through an acoustic lens to produce an acoustic image. An acoustic lens focuses sound in much the same way as an optical lens focuses light. Lenses based on refraction have been widely used to focus sound. Wehr et al. [42] created a non-linear acoustic lens using chains of spheres, in order to let the sound waves travel through each chain to meet at a specific focal point and form increased relative amplitude. Candelas et al. [44] demonstrated that an acoustic lens can be built with a single subwavelength slit surrounded by a finite number of grooves.

There are also a number of studies that explore the use of an array of rigid cylinders in air (known as the sonic crystal) to make such refractive lenses. A sonic crystal is an artificial crystal composed of a periodic alignment of acoustic scattering materials imbedded in the uniform host material as shown in Figure 3 [49]. It is expected to have full band-gaps, in which any acoustic wave cannot propagate in the crystal. For example, Sanchez-Perez et al. [48] showed an insufficient band-gap of a two-dimensional periodic array of rigid cylinders in air. Miyashita et al. [49, 50, 51] observed the existence of a full bandgap for a sonic crystal made of a periodic array of acrylic cylinders in air, and further studied its application as various shapes of sonic wave-guides to directionally transmit sound wave with a relative small leakage. Cervera et al. [46] used a sonic crystal to build up refractive

acoustic lenses for airborne sound. He et al. [47] proposed a hybrid sonic crystal imaging devices to achieve multi-images from one-source input along with the designable image positions.

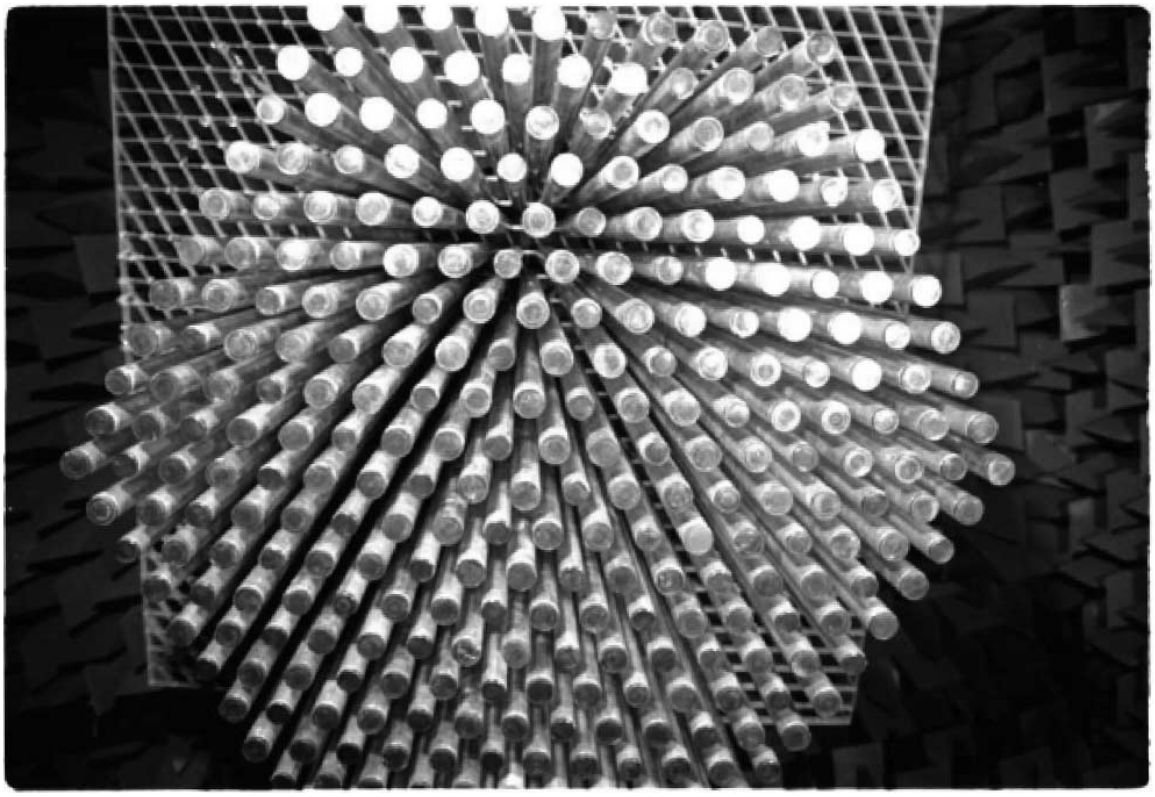

Figure 3 Photograph of a typical sonic crystal sample taken from underneath [48].

However, certain limitations exist to the aforementioned techniques. For lenses based on sonic crystals, the lattice constant must be of the same order as the acoustic wavelength, which results in an imaging resolution that is limited by diffraction. Moreover, as we are using a range of frequencies for the sound sources, it would be impractical to build such lenses for each of the testing frequency. Although new technologies such as acoustic diodes [16], acoustic rectifiers [17], and 3D acoustic metamaterials [45] have been

proposed recently to overcome these limitations, practical realization of such technologies has never been achieved.

2.2.2 Time-Reversal Strategy

In addition to physical lenses, previous works have described other strategies such as time-reversal techniques to focus sound and generate acoustic image. For instance, sound pressure can be increased in a specified space by array signal processing, such as beamforming [43], or side lobe suppression method. It is based on nullspace-based sound field control (NBSFC) [15], which suppresses the sound pressure at any control point with a loudspeaker array, and forms an area where only the desired sound is reproduced to each listener. Similar processing techniques have also been used on commercial directional loudspeaker products such as Acouspade [14], as well as 3D-printed speakers utilizing electrostatic loudspeaker technology [52, 53].

With sensor arrays, such time-reversal approach can be achieved by properly processing the received signals at each channel so that the target sound sources can be selectively amplified and imaged. Compared to the lens based strategy, time-reversal technique offers more flexibility in applications that require special manipulations of acoustic waves, such as medical imaging or therapy.

2.3 Array Signal Processing

Microphone arrays present additional challenges for signal processing methods since large numbers of detecting elements and sensors generate large amounts of data to be

processed. Furthermore, applications such as sound source localization and sound imaging, require complex algorithms to properly process the raw data [57]. On the one hand, the environmental noise in acoustical imaging is often severe, thus significant effort needs to be put into signal processing for the extraction of imaging data from noise. On the other hand, because of their low-propagation velocities, acoustic signals can be gated to provide range discrimination, resulting in a failure in the estimation of their locations. Therefore, many signal processing techniques that use arrays of sensors have been proposed to improve the quality of the output signal and achieve a substantial improvement in the signal-to-noise ratio (SNR) [3, 18, 56, 59]. The most well-known general class of array processing methods are beamforming [25], which has already been widely used for many decades in different application fields, such as the Sound Navigation and Ranging (SONAR), Radio Detection and Ranging (RADAR), telecommunications, and ultrasound imaging [26]. The beamforming technique requires the utilization of microphone arrays that capture all emanating sounds. All incoming signals are then combined to amplify the primary source signal, while at the same time suppressing any environmental noise.

2.3.1 Static Beamforming

Generally, there are two different types of beamforming: static (or non-adaptive) and adaptive [25]. Non-adaptive method is based on the fact that the spatial environment is already known and tracking devices are used to locate the sound sources. It involves using a fixed set of parameters for the transducer array, as the array processing parameters do not change dynamically over time. In the majority of the cases, a non-

adaptive delay-and-sum approach is utilized, due to its rather simple implementation and because a tracking device (such as a video camera) is almost always available. Algorithms such as Delay-and-Sum Beamforming (DSBF) [60] can be used to generate an acoustic map from microphone data. These algorithms generally use time delays which can be calculated from the geometry of the array.

In DSBF, the signals received by the microphones in the array are time-aligned with respect to each other in order to adjust for the path-length differences between the sound source and each of the microphones. The now time-aligned signals are then weighted and added together. Any interfering signals from noise sources that are not coincident with the sound source remain misaligned and are thus attenuated when the signals are combined. A natural extension to DSBF is filter-and-sum beamforming, in which each microphone has an associated filter, and the received signals are first filtered and then combined.

2.3.2 Adaptive Beamforming

In contrast, adaptive approaches do not utilize tracking devices to locate the sound source. In fact, the received signals from the microphones are used to calibrate properly the beamformer, in order to improve the quality of the extracted source. Adaptive beamforming can adapt the parameters of the array in accordance with changes in the application environment. These methods, such as the Minimum Variance Distortionless Response (MVDR) [57], Frost [61], and Generalized Sidelobe Canceller (GSC) [62], update the array parameters on a sample-by-sample or frame-by-frame basis according to

a specified criterion. Typical criteria used in adaptive beamforming includes a distortionless response in the look direction or the minimization of the energy from all directions not considered the look direction. Although computationally demanding, it can perform better than static beamforming in noise rejection.

The improved performance does not come without a price however. For example, the MVDR is more sensitive to sensor position errors. In circumstances where sensor positions are inaccurate, MVDR could produce a worse spatial spectrum than DSBF. Moreover, if the difference of two signal directions is further reduced to a level that is smaller than the beamwidth of an MVDR beam, the MVDR estimator will also fail. Consequently, the required microphones must be placed at very specific and carefully selected positions, which is unfeasible in many applications.

Adaptive beamforming is most useful only when the interference is concentrated in a small number of known directions or in some known set of frequencies. The algorithms are extremely sensitive to delay estimates, signal attenuation, and filtering characteristics, for example, enough to make them ineffective for many practical applications. Thus, for sound source localization and imaging, the static beamforming algorithm performed in the frequency domain is normally adopted as the fundamental processing method [41].

3 Chapter: Methodology

The most fundamental step in obtaining the source-origin information is estimating the time difference of arrival (TDOA) between different microphones. This estimation problem would be an easy task if the received signals were merely a delayed and scaled version of each other. In reality, however, the source signal is generally immersed in ambient noise since we are living in a natural environment where the existence of noise is inevitable. Furthermore, each observation signal may contain multiple attenuated and delayed replicas of the source signal due to reflections from boundaries and objects. This multipath propagation effect (termed as reverberation) introduces echoes and spectral distortions into the observation signal, which severely deteriorates the source signal. In addition, the source may also move from time to time, resulting in a changing time delay. All these factors make TDOA estimation a complicated and challenging problem.

As discussed in the previous chapters, beamforming technology alleviates the majority of the shortcomings that other recording techniques introduce, at the cost of an increased number of input channels. A beamformer is a processor used in conjunction with an array of sensors to provide a versatile form of spatial filtering [25]. The sensor array collects spatial samples of propagating wave fields, which are processed by the beamformer. The objective is to estimate the signal arriving from a desired direction in the presence of noise and interfering signals. The beamformer performs spatial filtering to separate signals that have overlapping frequency content but originate from different spatial

locations. The spatial-filter based beamformer was developed for narrowband signals that can be sufficiently characterized by a single frequency. It can be used for plenty of different purposes, such as detecting the presence of a signal, estimating the direction of arrival (DOA), and enhancing a desired signal from its measurements corrupted by noise, competing sources, and reverberation.

Currently, we are using a delay-and-sum beamformer, which is the simplest way of computing the beam. Delay-and-sum beamforming (DSBF) uses the fact that the delay for the sound wave to propagate from one microphone in the array to the next can be empirically measured or calculated from the array geometry. This delay is different for each direction of sound propagation, i.e., from the sound source position. By delaying the signal from each microphone by an amount of time corresponding to the direction of propagation and then summing the delayed signals, we selectively amplify the sound coming from a particular direction. This process is illustrated in Figure 4. DSBF assumes that the position of the desired source relative to the array is known. The problem of accurately localizing a source is crucial, but rather separate from the problem of amplifying sound coming from a particular direction.

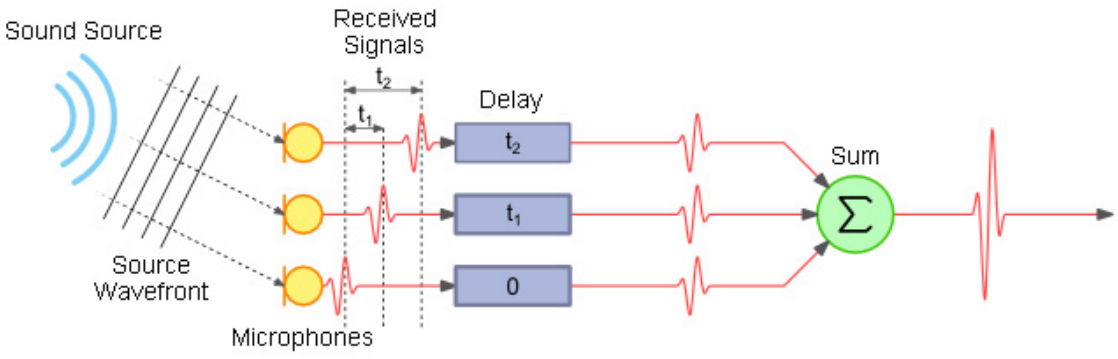

Figure 4 Illustration of the delay-and-sum beamforming process.

3.1 Speed of Sound

The sound is a compression wave which travels around 343.2 meters per second in dry air with temperature at 20 ˚C. The speed of sound in various temperature conditions can be calculated from the equation below [19]:

$$c = 331.3 \times \sqrt{1 + \frac{T}{273.15}} \quad (1)$$

where c is the speed of sound in m/s, and T is the temperature in ˚C.

The propagation speed of sound is significantly slower compared to light and electromagnetic waves, which lowers the demand for precise clock in a sound based localization system. For example, in a GPS receiver a clock error of one nanosecond introduces a range measurement error of 0.3 meter, whereas in a sound based localization system the same error will be introduced with a clock error of one millisecond.

3.2 Far-field and Near-field Cases

The wave front from a sound source is curved and not flat, which introduces much more complexity in the actual position calculations using TDOA. However, this wave front can be approximated as flat but it introduces an error in the calculated position. The relative size of this error compared to other error sources depends on the distance between the sound source and the microphone array and the size of the array. If the distance between the sound source and the array is large compared to the dimensions of the array the approximation does not introduce much additional error in the calculated position. This is

often referred to as a far-field case. If the distance between the sound source and the array is small compared to the dimensions of the array this approximation cannot be made. This is referred to as a near-field case.

In most beamforming applications two assumptions simplify the analysis: the signal sources are located far enough away from the array that the wave fronts impinging on the array can be regarded as plane waves (far-field assumption), and the signals incident on the array are narrow-banded (narrow-band assumption).

3.3 Problem Description

In sensor arrays, a widely used signal model assumes that each propagation channel introduces only some delay and attenuation. With this assumption and in the scenario where we have an array consisting of *N* sensors, the array outputs, at time *k*, can be expressed as:

$$y_n(k) = a_n s[k - t - \mathrm{F}_n(\tau)] + v_n(k) \tag{2}$$
$$= x_n(k) + v_n(k), n = 1,2, \dots, N,$$

where a_n $(n = 1,2, \dots, N)$, which range between 0 and 1, are the attenuation factors due to propagation effects, $s(k)$ is the unknown source signal, *t* is the propagation time from the unknown source to sensor 1, $v_n(k)$ is an additive noise signal at the nth sensor, τ is the relative delay (or TDOA) between sensor 1 and 2, and $\mathrm{F}_n(\tau)$ is the relative delay between sensors 1 and n with $\mathrm{F}_1(\tau) = 0$ and $\mathrm{F}_2(\tau) = \tau$. In this chapter, we make a key assumption that τ and $\mathrm{F}_n(\tau)$ are known or can be estimated and the source and noise

signals are uncorrelated. We also assume that all the signals in equation (2) are zero-mean and stationary.

By processing the array observations $y_n(k)$, we can acquire much useful information about the source, such as its position, frequency, etc. However, the problem considered in this chapter is focused on reducing the effect that the additive noise terms $v_n(k)$ may have on the desired signal, thereby improving the signal-to-noise ratio (SNR). Considering the first sensor as the reference signal, the goal of this chapter can be described as to recover $x_1(k) = a_1 s(k - t)$ up to an eventual delay.

3.4 Delay-and-Sum Beamforming

Instead of physically, beamforming algorithmically steers the sensors in the array toward a target signal. The direction the array is steered is called the look direction. In order to simulate the directivity of a microphone array, we need to make the following assumptions [55]:

- All microphones are identical and have unity gain and induce zero phase shifts to the recorded signal.
- Microphones are dot-like and do not alter the sound field. They individually have perfect spherical directivity.
- The impinging sound waves are plane waves.

The delay-and-sum (DS) beamformer consists of two basic processing steps [57, 70, 78]. The first step is to time-shift each sensor signal by a value corresponding to the TDOA

between that sensor and the reference one. With the signal model given in equation (2) and after time shifting, we obtain

$$\begin{aligned} y_{a,n}(k) &= y_n[k + \mathrm{F}_n(\tau)] \\ &= a_n s(k-t) + v_{a,n}(k) \\ &= x_{a,n}(k) + v_{a,n}(k), n = 1,2,\dots,N, \end{aligned} \tag{3}$$

where

$$v_{a,n}(k) = v_n[k + \mathrm{F}_n(\tau)],$$

and the subscript 'a' implies an aligned copy of the sensor signal. The second step consists of adding up the time-shifted signals, giving the output of the DS beamformer:

$$\begin{aligned} z_{DS}(k) &= \frac{1}{N}\sum_{n=1}^{N} y_{a,n}(k) \\ &= a_s s(k-t) + \frac{1}{N} v_s(k) \end{aligned} \tag{4}$$

where

$$\alpha_s = \frac{1}{N}\sum_{n=1}^{N} \alpha_n$$

$$v_s(k) = \sum_{n=1}^{N} v_{a,n}(k)$$

$$= \sum_{n=1}^{N} v_n[k + \mathrm{F}_n(\tau)]$$

In this way the DS beamformer is able to determine the amplitude of incident sound as a function of its frequency and direction of arrival.

3.5 Signal-to-Noise Ratio

Now we can examine the input and output SNRs of the DS beamformer [57]. For the signal model given in equation (2), the input SNR relatively to the reference signal is

$$SNR = \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2} = \alpha_1^2 \frac{\sigma_s^2}{\sigma_{v_1}^2} \tag{5}$$

where $\sigma_{x_1}^2 = E[x_1^2(k)]$, $\sigma_{v_1}^2 = E[v_1^2(k)]$, and $\sigma_s^2 = E[s^2(k)]$ are the variances of the signals $x_1(k)$, $v_1(k)$, and $s(k)$, respectively. After DS processing, the output SNR can be expressed as the ratio of the variances of the first and second terms in the right-hand side of equation (4):

$$\begin{aligned} oSNR &= N^2 \alpha_s^2 \frac{E[s^2(k-t)]}{E[v_s^2(k)]} \\ &= N^2 \alpha_s^2 \frac{\sigma_s^2}{\sigma_{v_s}^2} \\ &= \left(\sum_{n=1}^{N} \alpha_n \right)^2 \frac{\sigma_s^2}{\sigma_{v_s}^2} \end{aligned} \tag{6}$$

where

$$\begin{aligned} \sigma_{v_s}^2 &= E\left\{ \left[\sum_{n=1}^{N} v_n[k + \mathrm{F}_n(\tau)] \right]^2 \right\} \\ &= \sum_{n=1}^{N} \sigma_{v_n}^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \varrho_{v_i v_j} \end{aligned} \tag{7}$$

with $\sigma_{v_n}^2 = E[v_n^2(k)]$ being the variance of the noise signal $v_n(k)$, and $\varrho_{v_i v_j} = E\{v_i[k + \mathrm{F}_i(\tau)] v_j[k + \mathrm{F}_j(\tau)]\}$ being the cross-correlation function between $v_i(k)$ and $v_j(k)$.

Let us assume that the noise signals at the microphones are uncorrelated, i.e., $\varrho_{v_i v_j} = 0, \forall i, j = 1,2, \dots, N, i \neq j$, and they all have the same variance, i.e., $\sigma_{v_1}^2 = \sigma_{v_2}^2 = \cdots = \sigma_{v_n}^2$. We also suppose that all the attenuation factors are equal to 1 (i.e., $\alpha_n = 1, \forall n$). Then it can be easily checked that

$$oSNR = N \cdot SNR \tag{8}$$

It is interesting to see that under the previous conditions, a simple time-shifting and adding operation among the sensor outputs results in an improvement in the SNR by a factor equal to the number of sensors. This will mean that the signal $z_{DS}(k)$ will be less noisy than any microphone output signal $y_n(k)$, and will possibly be a good approximation of $x_1(k)$.

3.6 Beam Pattern

Another way of illustrating the performance of a DS beamformer is through examining the corresponding beam pattern [57], which provides a complete characterization of the array system's input-output behavior. The term beam pattern characterizes the array's input-output behavior when the beamformer is steered to a specific direction. It can be used to analyze how the array output is affected by signals different from the focused one.

From the previous analysis, we easily see that a DS beamformer is indeed an N-point spatial filter and its beam pattern is defined as the magnitude of the spatial filter's directional response. From equation (3) and (4), we can check that the nth coefficient of

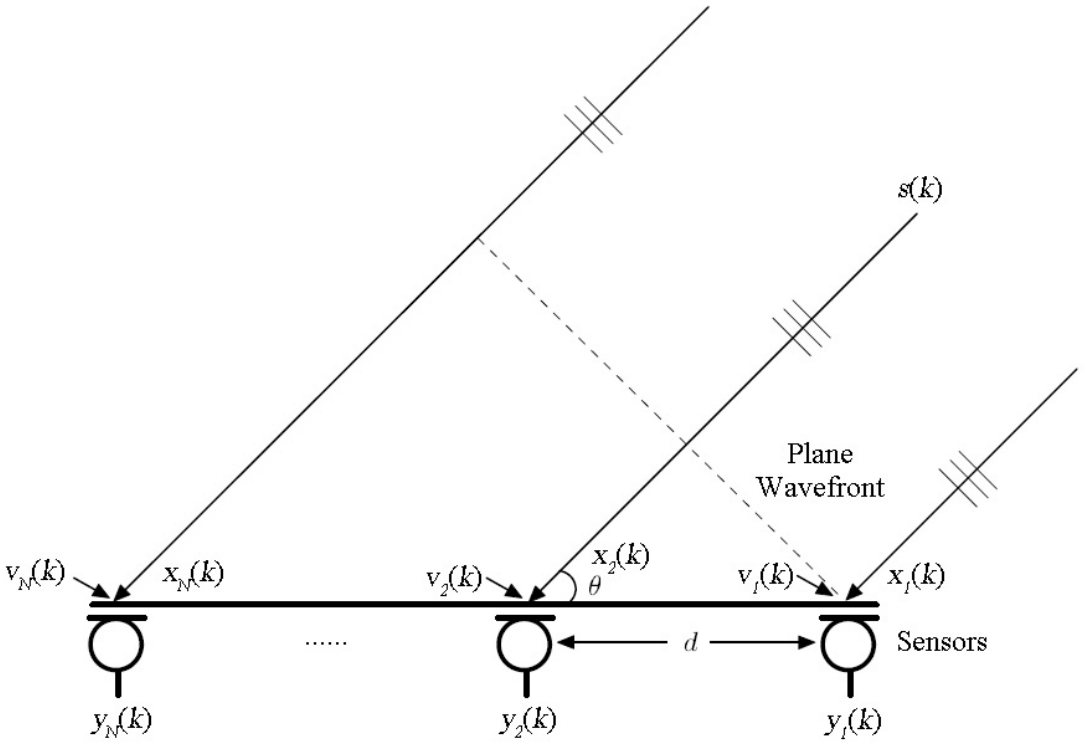

Figure 5 Illustration of an equispaced linear array, where the source *s(k)* is located in the far field, the incident angle is θ, and the spacing between two neighboring sensors is *d*.

the spatial filter is $\frac{1}{N} e^{j2\pi f \mathrm{F}_n(\tau)}$, where *f* denotes frequency. The directional response of this filter can be found by performing the Fourier transform. Since $\mathrm{F}_n(\tau)$ depends on both the array geometry and the source position, so the beam pattern of a DS beamformer should be a function of the array geometry and source position. In addition, the beam pattern is also a function of the number of sensors and the signal frequency. Now suppose that we have an equispaced linear array, which consists of *N* omnidirectional sensors, as illustrated in Figure 5. If we denote the spacing between two neighboring sensors as *d*, and assume that the source is in the far field and the wave rays reach the array with an incident angle of θ, the TDOA between the nth and the reference sensors can be written as

$$\mathrm{F}_n(\tau) = (n-1)\tau = (n-1)\frac{d\cos\theta}{c} \tag{9}$$

where c denotes the sound velocity in air, and can be calculated from equation (1). In this case, the directional response of the DS filter, which is the spatial Fourier transform of the filter [80], can be expressed as

$$\begin{aligned} S_{DS}(\varphi,\theta) &= \frac{1}{N}\sum_{n=1}^{N}[e^{j2\pi(n-1)fd\cos\theta/c}]e^{-j2\pi(n-1)fd\cos\varphi/c} \\ &= \frac{1}{N}\sum_{n=1}^{N}e^{-j2\pi(n-1)fd\,[\cos\varphi-\cos\theta]/c} \end{aligned} \tag{10}$$

where φ $(0 \le \varphi \le \pi)$ is a directional angle. The beam pattern is then written as

$$\begin{aligned} A_{DS}(\varphi,\theta) &= |S_{DS}(\varphi,\theta)| \\ &= \left|\frac{sin[N\pi fd\,(cos\,\varphi - cos\,\theta)\,/c]}{Nsin[\pi fd\,(cos\,\varphi - cos\,\theta)\,/c]}\right| \end{aligned} \tag{11}$$

Using the phased array system toolbox in MATLAB [68], we simulated the beam pattern for an equispaced linear array with ten sensors, d = 8 cm, θ = 90°, and f = 2 kHz. Figure 6 plots the result. It consists of a total of 9 beams (in general, the number of beams in the range between 0° and 180° is equal to N − 1). The one with the highest amplitude is called mainlobe and all the others are called sidelobes. One important parameter regarding the mainlobe is the beamwidth (mainlobe width), which is defined as the region between the first zero-crosses on either side of the mainlobe. With the above linear array, the beamwidth can be easily calculated as $2\cos^{-1}[c/(Ndf)]$. This number decreases with the increase of the number of sensors, the spacing between neighboring sensors, and the signal frequency. The height of the sidelobes represents the gain pattern for noise and

competing sources present along the directions other than the desired look direction. In array and beamforming design, we hope to make the sidelobes as low as possible so that signals coming from directions other than the look direction would be attenuated as much as possible.

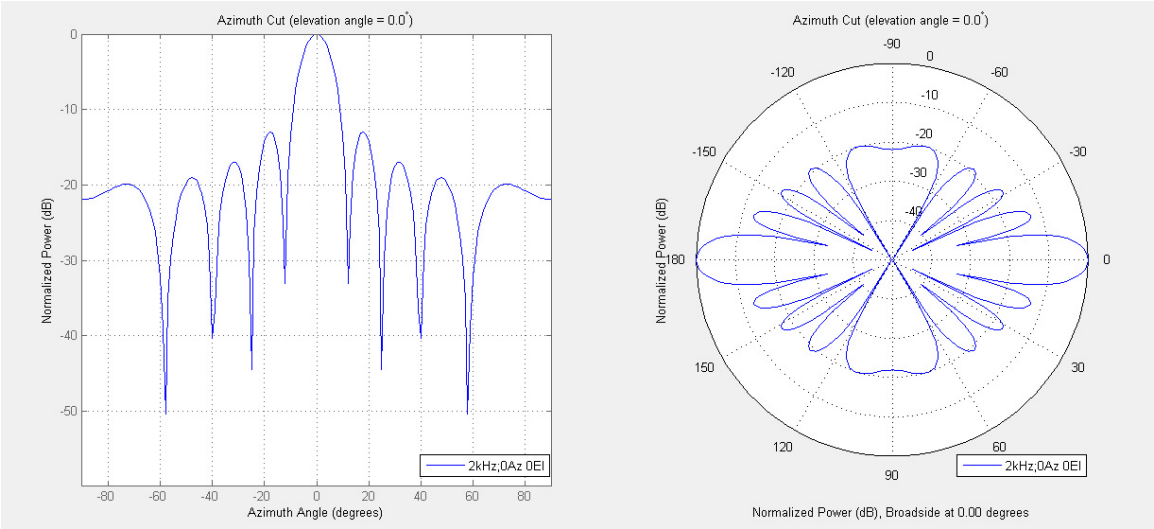

Figure 6 Beam pattern of a ten-sensor array when $\theta = 90°$, $d = 8$ cm, and $f = 2$ kHz: in Cartesian coordinates (left) and in polar coordinates (right).

3.7 Sensor Spacing

From the previous analysis, we see that the array beamwidth decreases as the spacing d increases. So, if we want a sharper beam, we can simply increase the spacing d, which leads to a larger array aperture. This would, in general, lead to more noise reduction. Therefore, in array design, we would expect to set the spacing as large as possible. However, when d is larger than $\lambda/2 = c/(2f)$, where λ is the wavelength of the signal, spatial aliasing would arise [57]. To visualize this problem, we plot the beam pattern for an equispaced linear array same as used in Figure 6 (right). The signal frequency f is still

2 kHz. But this time, the array spacing d is 24 cm. The corresponding simulation result is shown in Figure 7. This time, we see three large beams that have a maximum amplitude of 1. The other two are called grating lobes. Signals propagating from directions at which grating lobes occur would be indistinguishable from signals propagating from the mainlobe direction. This ambiguity is often referred to as spatial aliasing. In order to avoid spatial aliasing, the array spacing has to satisfy

$$d \leq \frac{\lambda}{2} = \frac{c}{2f} \tag{12}$$

By analogy to the Nyquist sampling theorem, this result may be interpreted as a spatial sampling theorem.

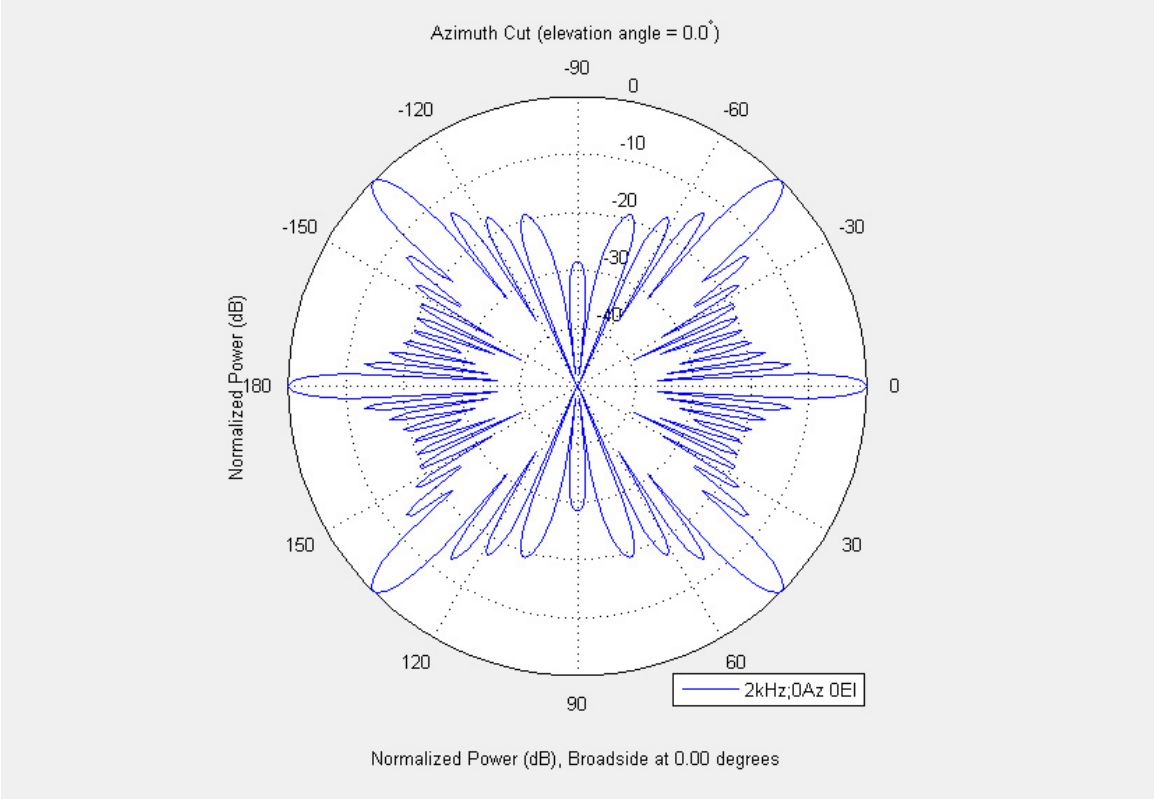

Figure 7 Beam pattern (in polar coordinates) of a ten-sensor array when $\theta = 90°$, $d = 24$ cm, and $f = 2$ kHz.

The above case suggests two requirements in array design: the spacing among sensors cannot be too large (as compared to the wavelength). Otherwise we will experience the spatial aliasing problem, which causes ambiguity in recovering the desired signal. On the other hand, the sensors cannot be too close. If they are too close, the array does not provide enough aperture for recovering the source signal. A general rule of thumb is to choose the spacing of sensors between $\lambda/10$ and $\lambda/2$ [81, 82].

4 Chapter: System Design

The design of a beamforming system requires various evaluations before its final implementation. Based on the applications, the size of the recording area and the hardware cost limitations, we need to evaluate different array geometries and the SNR quality of the extracted sources under different numbers of microphones. Furthermore, internal signal calculations, except filtering, in many cases also require decimation and interpolation. Based on the available hardware resources, we should carefully evaluate the sampling rate and the filtering coefficients of the algorithm used.

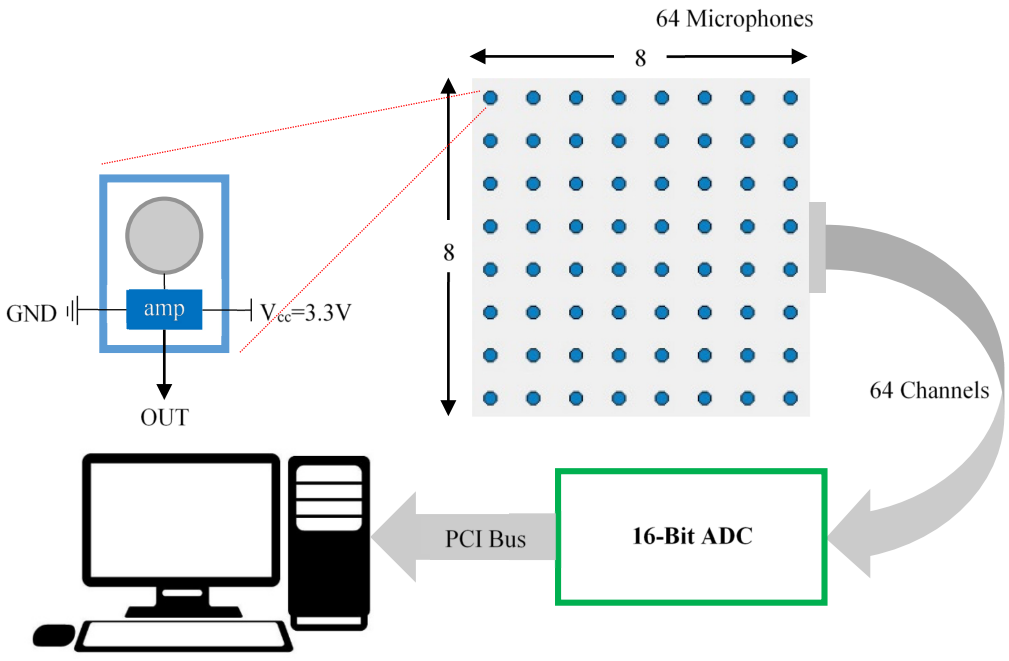

Figure 8 The schematic diagram of the 64-node microphone array hardware design.

4.1 Hardware Design

The design of our microphone array faces several challenges such as number of detectors, array geometry, reverberation, interference issues and signal processing. These factors are crucial to the construction of a reliable and effective microphone array system. Figure 8 illustrates the hardware design of the array.

4.1.1 The Sensors

Our array is constructed using the microphone modules from Adafruit [83]. The module consists of an electret condenser microphone (CUI CMA-4544PF-W) and an operational amplifier (Maxim MAX4466) with adjustable gain. Figure 9 shows the selected microphone module. The microphone is omni-directional and has a sensitivity of -44 ±2 dB at 1 kHz, 1 Pa. Its frequency response is essentially flat from 50 to 3000 Hz (Figure 9 right), and its operating frequency ranges from 20 Hz to 20 kHz. These electret microphones are fundamental parts of the array due to their small size, high sensitivity, and low power consumption. When combined into an array, they provide a highly versatile acoustic aperture.

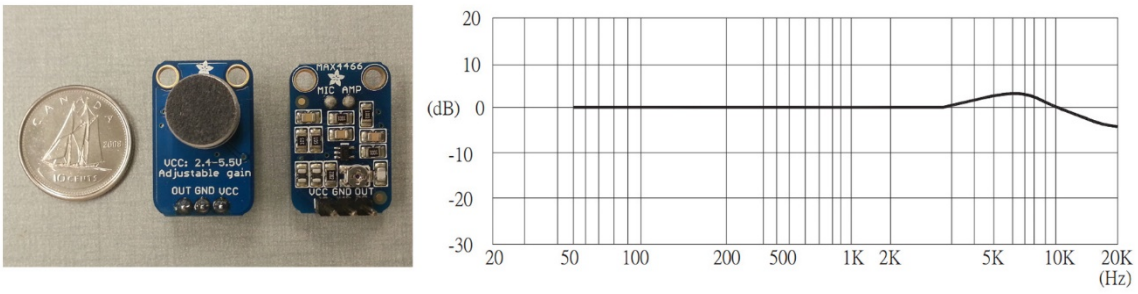

Figure 9 The microphone module selected for our array (left) and its frequency response curve (right).

4.1.2 Array Geometry

Depending on the nature of the applications, the geometry of the microphone array can play an important role in the formulation of the processing algorithms. For example, in sound source localization the array geometry must be known in order to be able to localize a source properly. Moreover, sometimes a regular geometry will even simplify the problem of estimation, which is why uniform linear, rectangular and circular arrays are often used [57]. Nowadays, these three geometries dominate the research but we see more and more sophisticated three-dimensional spherical arrays as they can better capture the sound field [34, 39]. However, in some other crucial problems such as noise reduction or source separation, the geometry of the array may have little (or no) importance depending on the algorithm.

Many array geometries have been suggested in past works [57], from linear to rectangular to circular; and similarly, many microphone spacing schemes have been suggested, from uniform to logarithmic. Different applications require different geometries in order to achieve optimum performance. In general, the disposition patterns of microphone array can be divided into linear array, planar array and three-dimensional array. Linear arrays are usually applied in medical ultrasonography, planar arrays are often used in sound source localization, and three dimensional spherical arrays are most frequently used in sophisticated SONAR applications. The size of an array is usually determined by what frequency the array will operate and what kind of spatial resolution the application using the acoustic array requires [31].

As mentioned in Chapter 3.7, the narrowband beamformer suffers from spatial aliasing at high frequencies. Spatial aliasing happens if a sensor spacing is wider than half of the signal wavelength. It is analogous to temporal aliasing in discrete-time signal processing. This leads to the condition that the spacing between any pair of microphones in the array should not exceed half of the smallest wavelength present in the signal. Spatial aliasing can also be avoided when the array geometry is totally non-redundant, that is, no difference vector between any two sensor positions is repeated [20]. For non-redundant arrays, which typically have an irregular or random geometry, the sidelobes causing ghost-images are suppressed. In general, irregular arrays outperform regular array designs, but it is difficult to find out how the design should be modified to obtain the best performance. Therefore, regular geometries where sensors are evenly spaced are preferred in order to simplify the algorithm development.

The performance of a microphone array typically increases linearly with the size of the array. This is well established in the theoretical literature on microphone arrays [31, 57]. The number of microphones is a compromise between the size of the board, the electronics needed to deal with the massive amount of data, and the need to achieve a relatively high SNR. Figure 10 gives the approximate peak SNRs and the recognition accuracies from the LOUD microphone array [31, 76], displaying the trend of improvement as the number of microphones is increased. The most drastic jump in the recognition accuracy curve can be seen when the number of microphones goes from 32 to 60, and after this point, adding more microphones does not make the curve increase faster.

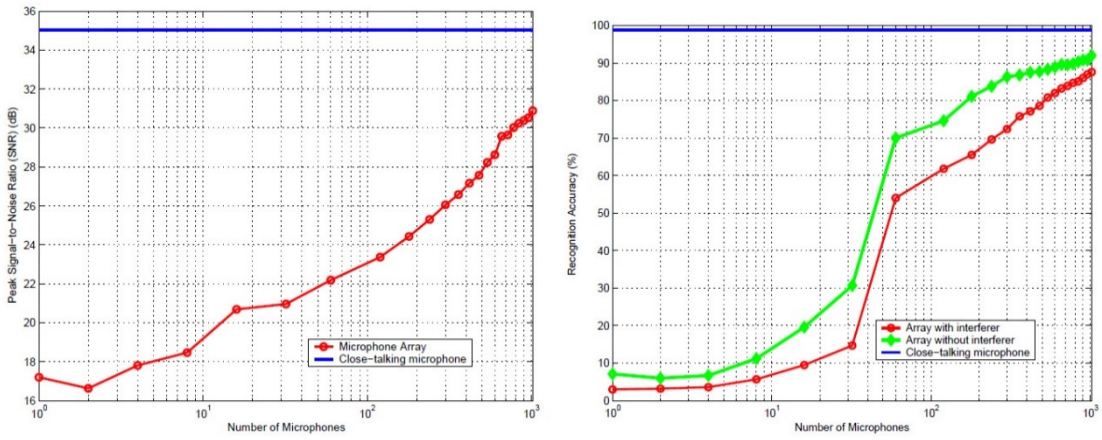

Figure 10 Experimental results from the LOUD microphone array [31, 76]: peak SNRs for one representative recording (left), and experimental recognition accuracies (right).

While many geometrical configurations of the array are possible and potentially desirable, our microphone array (pictured in Figure 14) consists of 64 elements, and is laid out in a rectangular grid pattern, with 8 columns and 8 rows. This allows us to steer the amplification beam horizontally as well as vertically. The schematic diagram of the layout is shown in Figure 8. In order to avoid spatial aliasing and have a good acoustic aperture, the inter-microphone distance is determined to be 2.3 cm center to center, and is maintained in both the horizontal and vertical directions. This decision was due both to practicality reasons, as well as preliminary experiments with various spacings. The 2.3 cm spacing is approximated by dividing the sound speed by the highest signal frequency and then dividing the result by two in order to satisfy the Nyquist-Shannon sampling theorem. This spacing makes it possible to obtain relevant phase information of incoming acoustic sound waves, increasing the array sensitivity and allowing spatial sampling of frequencies up to 7400 Hz without aliasing.

4.1.3 Data Acquisition System

A data acquisition system is necessary to condition, process, store and display the signals received by the array. Data acquisition for an acoustic array can be very challenging, but most of the issues are co-related with the type of sensors, number of sensors and array geometry. Also, the application plays an important role in defining the needs of the array for signal processing or real-time operation requirements.

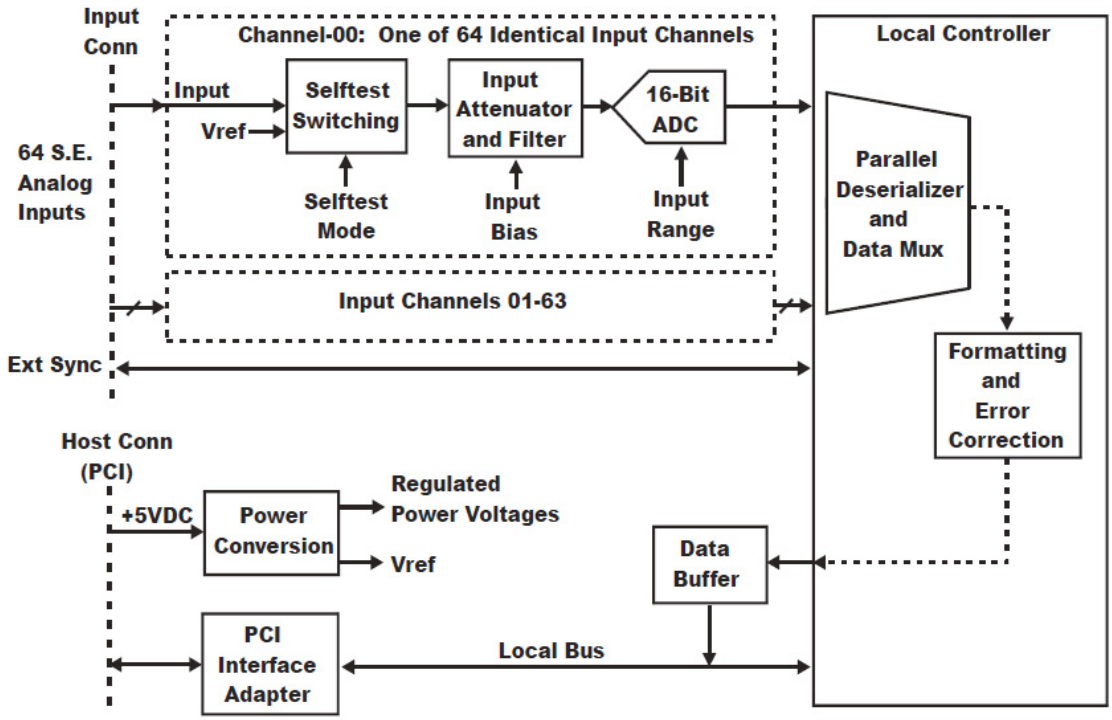

Figure 11 Functional block diagram of the PMC66-16AI64SSA data acquisition board.

To sample all the 64 microphone outputs simultaneously, we used the PMC66-16AI64SSA board from General Standards [67] for the data acquisition. This analog input board is a single-width PCI mezzanine card (PMC) that samples and digitizes 64 single-ended input channels simultaneously at rates up to 200,000 samples per second for each channel. It contains a dedicated 16-Bit sampling ADC for each input channel, and a

512K-Sample FIFO data buffer. As illustrated in Figure 11, serial data from each ADC is deserialized and multiplexed into a parallel data stream within the local controller. The output of the data multiplexer passes through a digital processor which applies gain and offset correction values obtained during auto-calibration. The corrected 16-bit data is formatted as offset binary, and then loaded into the FIFO data buffer. Finally, all the data is fed to the PCI bus through the interface adapter, and saved as a binary file in the host computer.

4.2 Software Design

In order to utilize the microphone array to selectively amplify sound coming from a particular direction, we applied a delay-and-sum beamforming (DSBF) algorithm as part of the processing software. Figure 12 shows the algorithms developed for our acoustic imaging system. These fall into four general areas:

- Importing the data,
- Constructing the delay-and-sum beamformer,
- Aiming the array and performing the beamforming,
- Filtering the results and generating the image.

The DSBF algorithm being used for this system was implemented in MATLAB [68]. The audio signal collected at each microphone in the array was streamed into the PMC input port. After the data acquisition, all of the signal data was saved as 16-bit offset binary, with a tag attached to all channel-00 data from the board. In order to import this data into MATLAB, we created a data logger which could locate every channel of data for each

time frame, transform the lower 16 bits into decimal, and store the reorganized data in a matrix with 64 columns, where each column represents the received signal at one of the array elements.

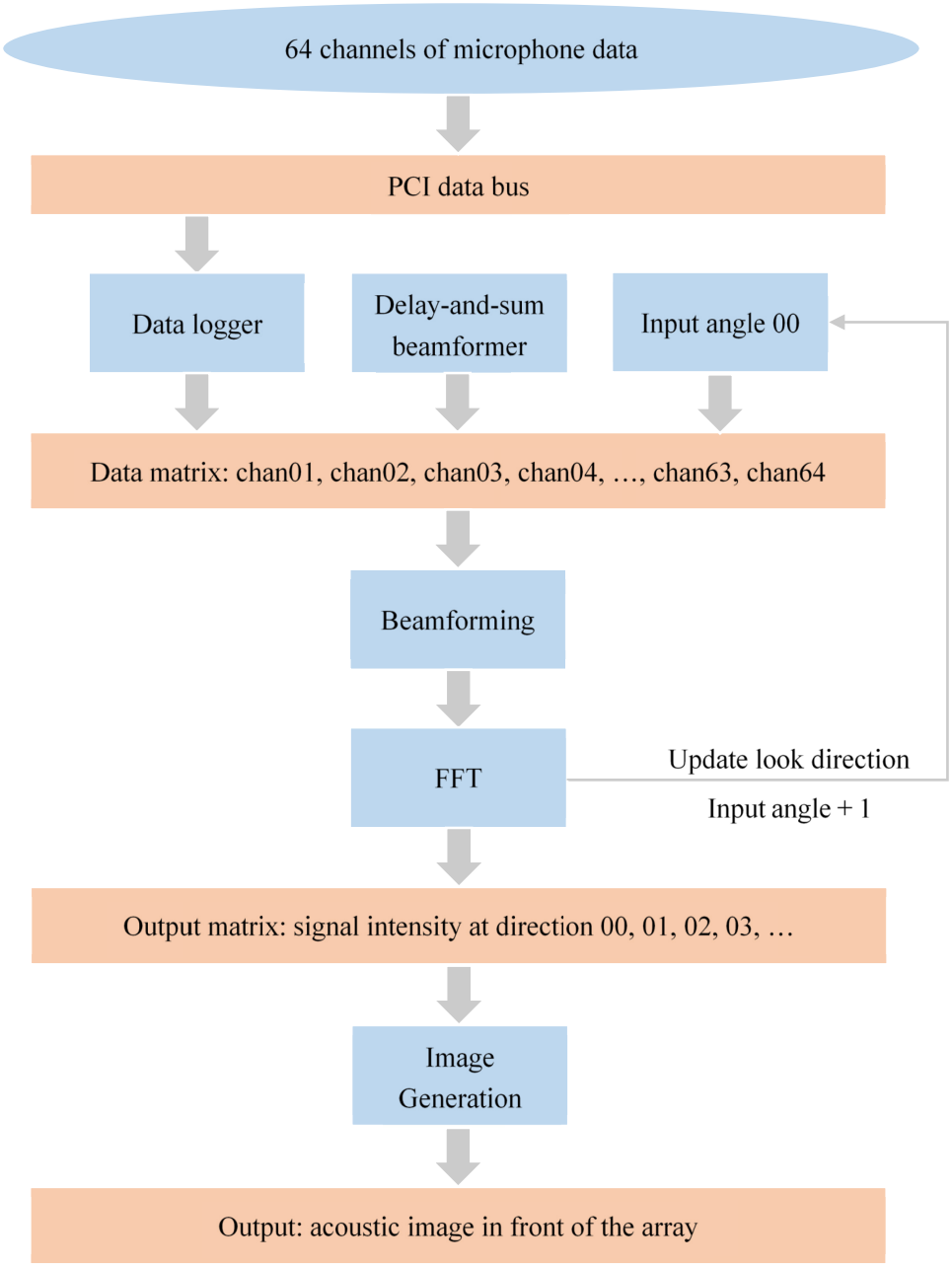

Figure 12 Block diagram of algorithms for our acoustic imaging microphone-array system.

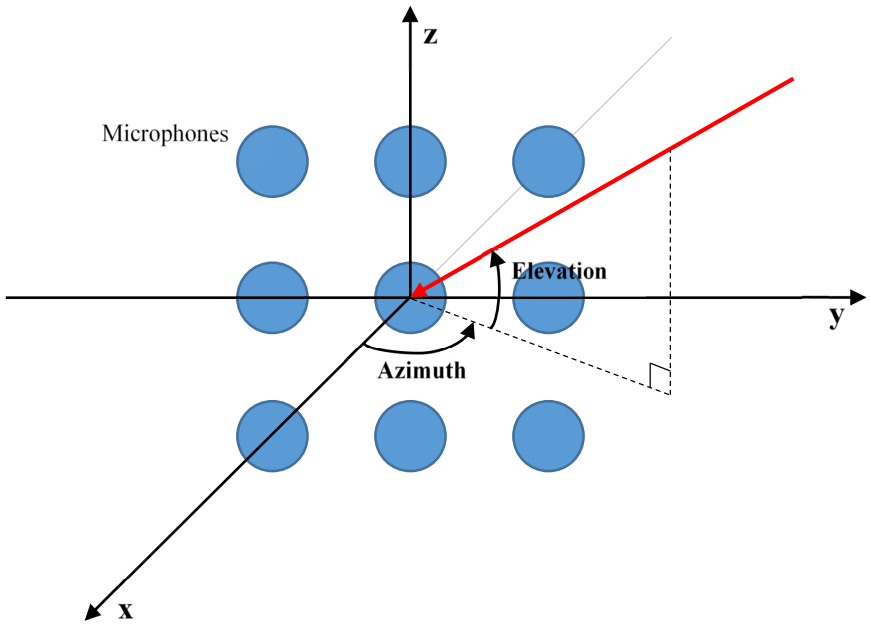

Figure 13 Azimuth and elevation angle with respect to the uniform rectangular array.

To perform the beamforming, we first defined a uniform rectangular array (URA) system object that stored the geometric information of the microphone array (such as size and element spacing). According to the array geometry, signal operating frequency and propagation speed, we then constructed the delay-and-sum beamformer and aimed it by appropriately delaying, attenuating, and adding the signals from each microphone so that they were in phase for one or more specific spatial locations. For applications such as speech enhancement, the location of the sound source needs to be known beforehand, in order to decide the look direction (input angle) of the beamformer. Since our microphone array is designed for the purpose of imaging, we are more interested in estimating the sound field at any point in front of the array, not just obtaining specific positions. Thus we need to utilize the beamformer to steer and scan the entire region in front of the array, which means the input angle needs to cover all 180 degrees in both azimuth (Az) and

elevation (El). The relationship between the two angles is illustrated in Figure 13. To achieve such purpose, we updated the input angle every time after the beamforming operation was performed so that it could scan the region from -90 degree to 90 degree in azimuth and elevation. In addition, we also performed a fast Fourier transform (FFT) on the beamformed signal so as to bandpass filter the result to further reduce the low-frequency noise, and extract the signal containing the frequency of the sound source to acquire its relative intensity level. Finally, with these signal intensities at every direction in front of the array, we were able to generate the acoustic image and reconstruct the sound field.

4.3 System Implementation

The body of the microphone array was 3D printed using currently available 3D printing technology. This allowed all the 64 microphones to be accurately positioned, maintaining a spacing of 2.3 cm in both the vertical and horizontal directions (Figure 14). For the best performance, all the microphones were powered from a 3.3V and 0.014A DC power supply, and all the output pins were connected to the PMC data acquisition board through a Mini D Ribbon (MDR) cable. After the audio data was transferred to the PCI bus, a computer connected to the PMC board was capable of retrieving the data using the 16AI64SSA driver application program interface (API) software. This software also contained an auto-calibration function that calibrated all analog input channels to a single internal voltage reference in order to obtain maximum measurement accuracy.

The system presented in this work was set to work at 50k samples/second which was well above the Nyquist rate to avoid aliasing. The data collection time was 1 second for all of the 64 channels. Once the sampling rate and time were set, the system went into data acquisition mode, where the boards were continuously sending data to the computer and the computer piped this data to a binary file on the hard disk. After the collection, we used the data logger to import this file so that we could visualize the raw data on the screen, and further process and analyze the data with the algorithms we developed for the system.

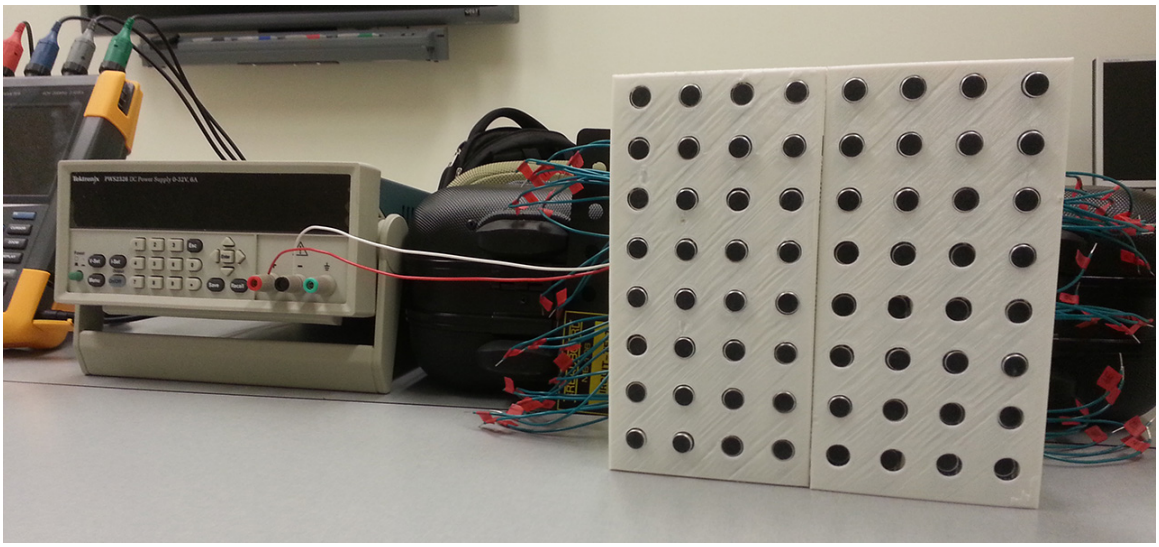

Figure 14 A photograph of the 64-node microphone array system.

Before conducting any experiment, it is necessary to calibrate the system. By adjusting the small trimmer pot on the back of each microphone module, we were able to calibrate the individual gain of each channel on the microphone array so that it could respond homogeneously when excited.

A particularly interesting use of this array is to steer it to various directions and create an intensity map of the acoustic power in various frequency bands via beamforming. The resulting image, as it is linked with direction, can be used to identify sound sources and relate them with physical objects in the world. We will discuss these applications in more details in Chapter 5 and 6.

4.4 Cost

Large microphone array setups are often too costly to design and deploy, such as those for small meeting spaces and auditoriums. In this work we set out to design a system that is relative inexpensive to build, install and maintain. The microphone modules we selected cost around $400 in total, and the price for the 3D prints is about $30. Depending on the applications, the cost for the data acquisition system may vary. The PMC board we used was a sample board requested from the General Standards Corporation which costs around $200. Given the aperture of our array (approximately 19×19 cm), we can calculate the cost-per-unit-area, which is \$630/361 cm^2 = \$1.745 sq cm, as our cost metric. Compared to commercial microphone array products and other research projects with relatively the same array size (whose cost-per-unit-area is usually above \$5 sq cm), our system costs roughly 65% less to build and deploy.

5 Chapter: Experiments and Results

5.1 Testing Environment

The experiment was carried out in a laboratory room with various random objects around, resulting in a noisy and highly reflective environment (as shown in Figure 15). The main noise sources were several cooling fans for computers and a loud air conditioner. All of the experiments were conducted under the same noise conditions. We did not setup any extra noise sources during the experiments. The room temperature was approximately 18 ˚C during the test sessions, from which we were able to estimate the speed of sound to be 342 m/s using equation (1).

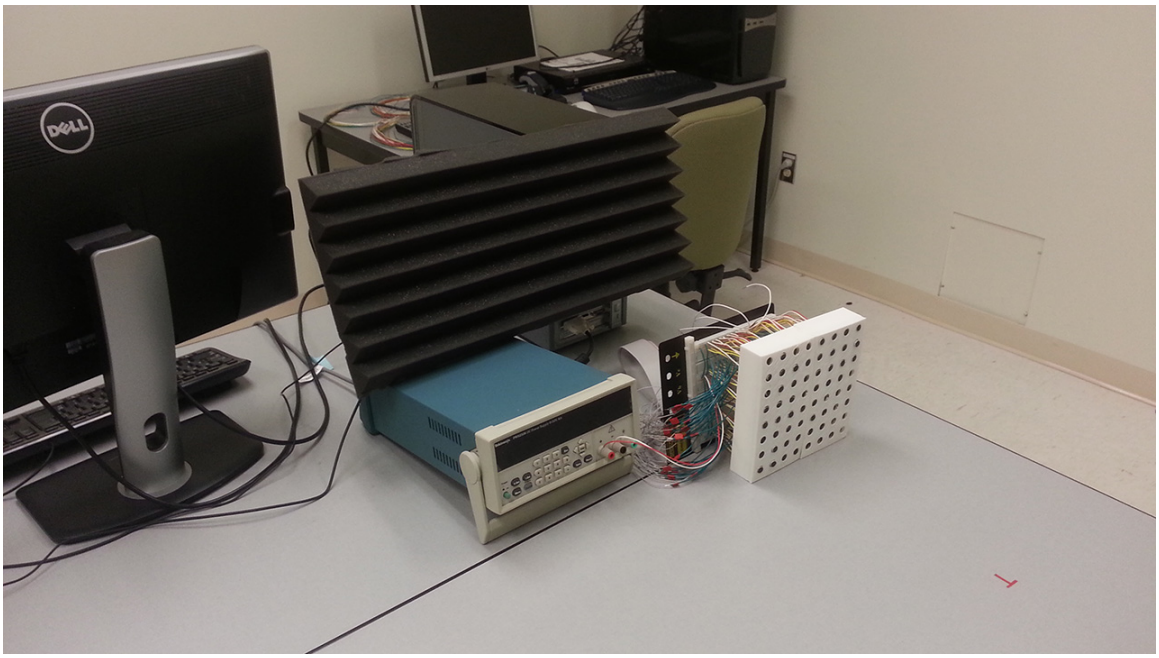

Figure 15 Testing environment for the microphone array beamforming system.

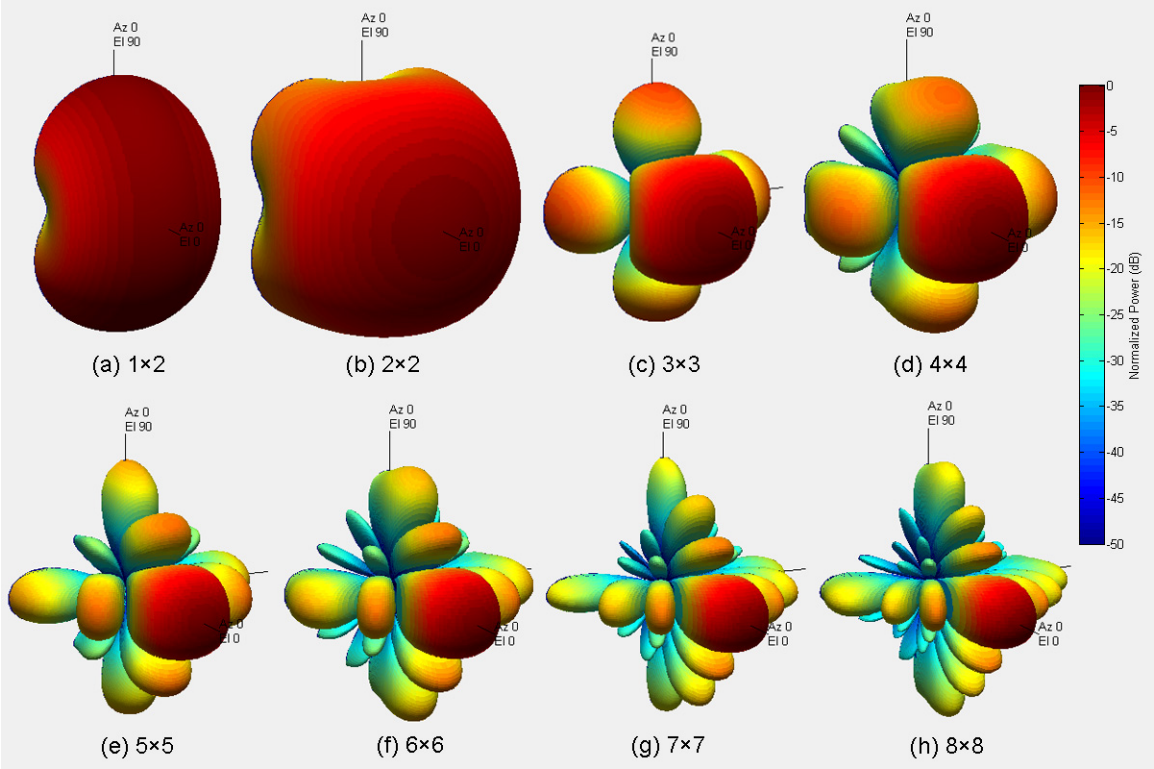

Figure 16 Beam patterns of rectangular arrays with different geometries: (a) 1 by 2, (b) 2 by 2, (c) 3 by 3, (d) 4 by 4, (e) 5 by 5, (f) 6 by 6, (g) 7 by 7, (h) 8 by 8.

5.2 Array Beam Pattern

As discussed in Chapter 3.6, once the array geometry is fixed and the desired steering direction is determined, the characteristics of the beam pattern of a DS beamformer, including the beamwidth, the amplitude of the sidelobes, and the positions of the nulls, would be fixed. We plot the beam patterns for microphone arrays with various geometries in Figure 16 when they are steered at 0 Az and 0 El. The sensor spacing is 2.3 cm, and the signal frequency is 7 kHz. As can be seen, with the increase of the array size, the beam pattern will have a narrower mainlobe which can provide more interference reduction. The number of lower sidelobes will also increase to bring more noise suppression. Figure 17 gives the array gains in dB for a look direction of 0 Az and 0 El,

displaying the trend of improvement as the number of microphones is increased. The array gain improves from 3.0 dB with two microphones to 18.1 dB with 64 microphones, which indicates a massive enhancement in SNR between the array output and each individual channel input. This result clearly demonstrates that having a large array size will significantly improve the quality of the beamformed signal.

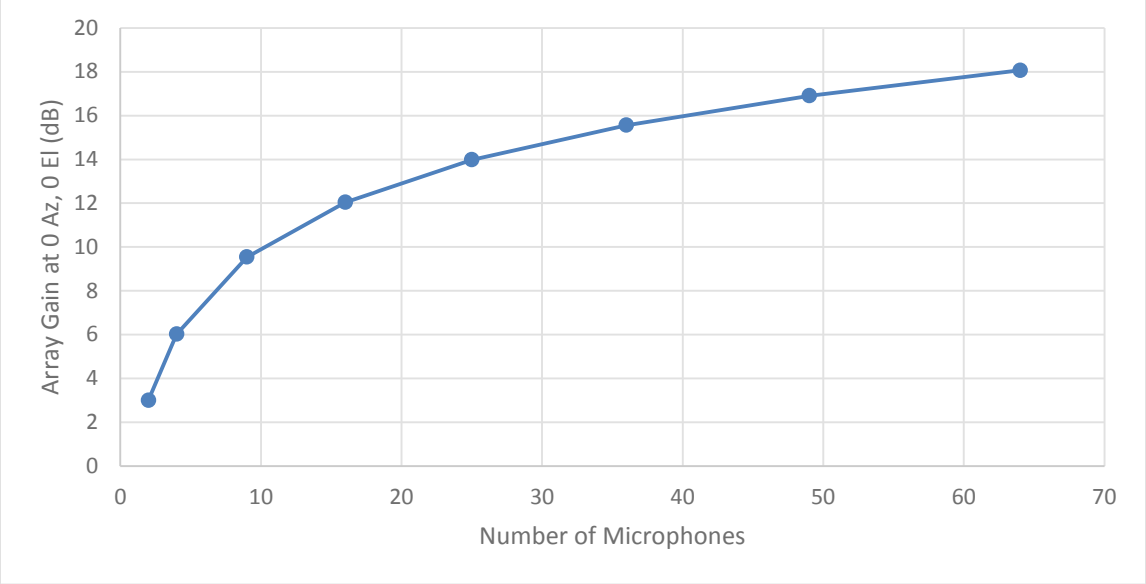

Figure 17 Array gains as the number of microphones is increased. The look direction is at 0 Az and 0 El.

5.3 Beamforming

A preliminary experiment was carried out to examine the effect of the DS beamformer. We used the microphone array to record a 5 kHz tone played in front of the array for 1 second in order to acquire sufficient signal samples. The recordings were sampled at 25 kHz, and then processed by the DSBF algorithm. A comparison between the raw data collected at one of the 64 channels (channel 35) and the overall beamformed signal is

shown in Figure 18 (left). The look direction of the beamformer was set to be -5 degrees in azimuth and 20 degrees in elevation. It can be seen from the result that after beamforming, the signal has produced a significant improvement over the raw data in terms of noise suppression, and has become a better approximation of the original 5 kHz sinusoidal signal. Furthermore, we also compare the frequency spectrum of the two signals in Figure 18 (right). In the spectrum of the channel 35 signal, we see that the low-frequency noise, especially at 64 Hz (1.64E+05) and 124 Hz (3.11E+05), constitutes a large portion of the signal power, which makes the 5 kHz source signal (2.46E+05) relatively weak. After beamforming, this 5 kHz component has been increased to 4.12E+05, which becomes much stronger compared to the noise, while the noise power at 64 Hz (1.09E+05) and 124 Hz (1.59E+05) has been significantly reduced along with other low-frequency components.

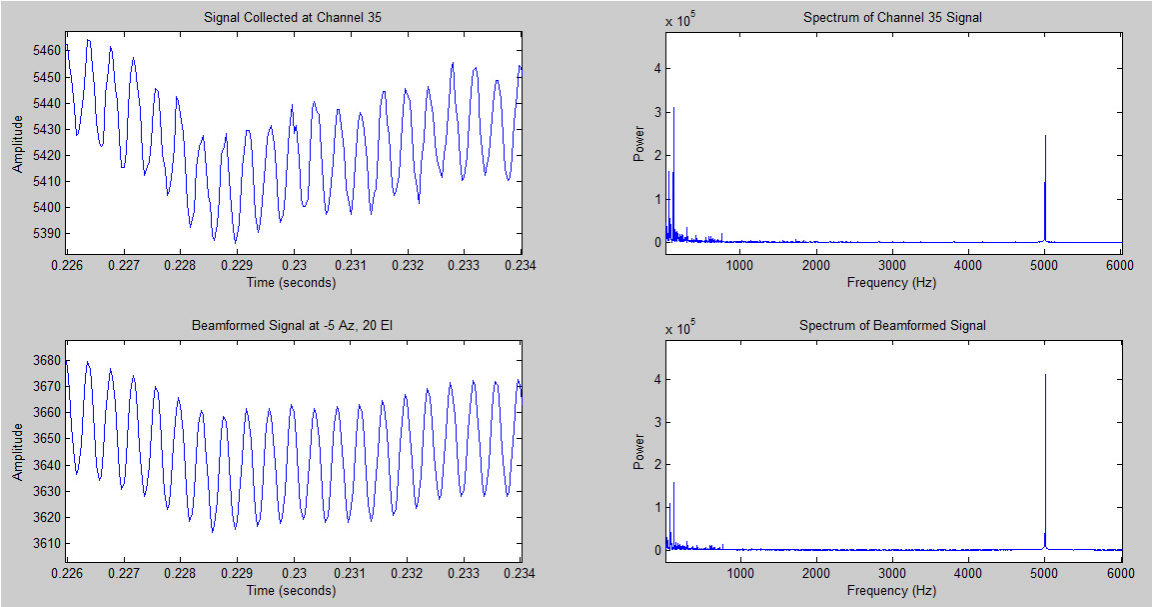

Figure 18 Comparison of the collected signal before and after beamforming (left), and their corresponding spectrums (right).

5.4 Source Localization and Separation

5.4.1 Sound Source Localization

To demonstrate its performance for source localization and separation, we have conducted several experiments with the 64-node array system. The localization experiments involved recording a moving sound source in a room where reverberation and several sources of noise were present. A wooden board with an 8×8 grid in the middle was placed 30 cm in front of the array (Figure 21). Each cell in the grid was 2.3 cm by 2.3 cm in order to match the spacing of the array. A speaker, 30 mm in diameter, producing continuous sinusoidal waves of 5 kHz was placed in one of the cells on the wooden board. This ensured that the speaker could be accurately positioned at every direction in front of the rectangular array. To prevent external noise from the computer cooling fan, sound-absorbing material was packed behind the testing area. The experiment setup is shown in Figure 19 top. The speaker was moved between every cell in the grid while the tone was recorded by the microphone array. Each recording was 1 second long and was sampled at 50 kHz. With the data collected at each 64 microphones, we were able to measure the relative sound pressure level at every direction in front of the array and visualize the intensity distribution of the sound field in order to localize the source signal.

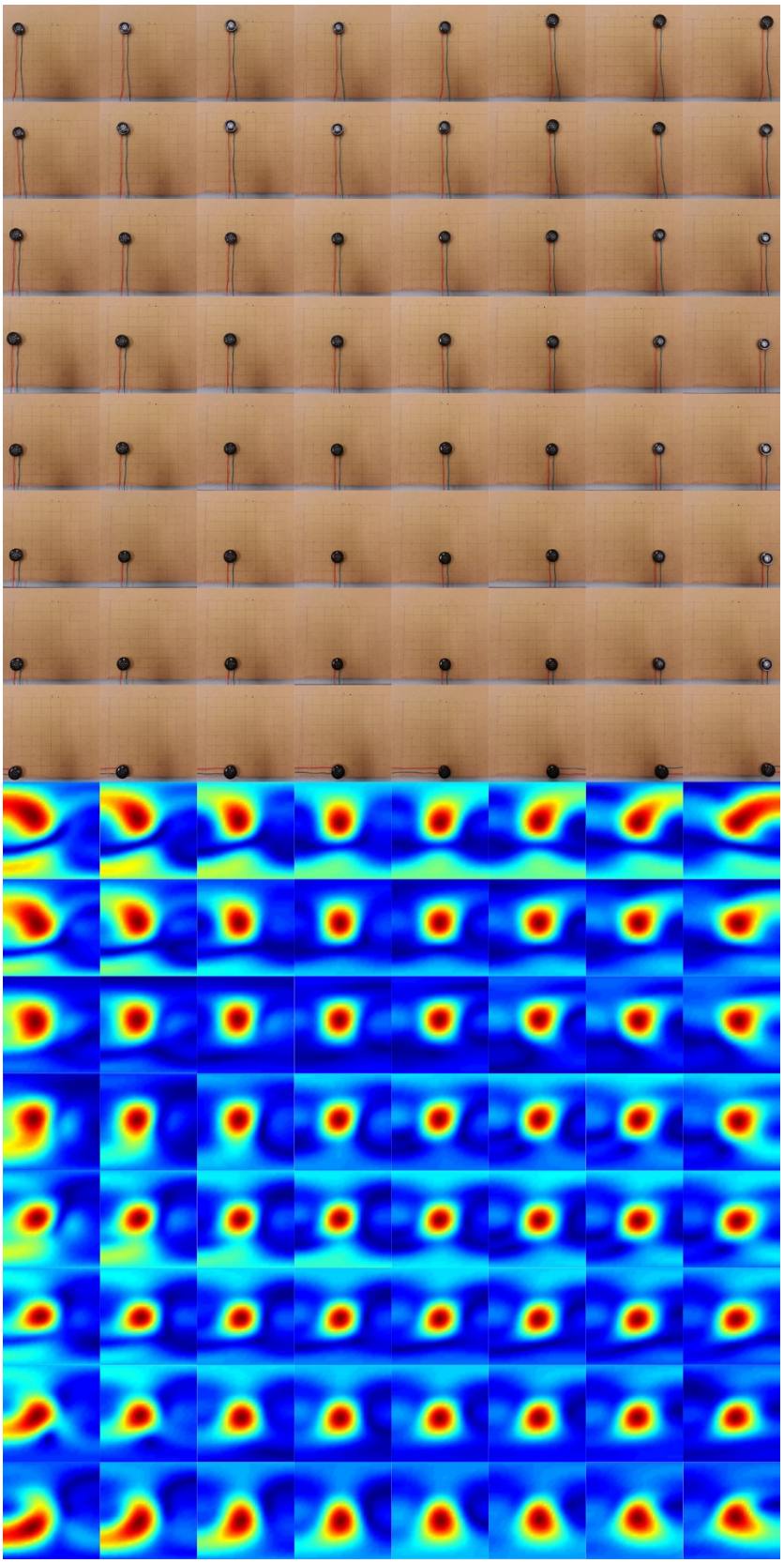

Figure 19 Setups for sound localization at 64 positions (top) and the corresponding results (bottom).

Figure 19 bottom displays the results of the sound source localization at 64 different positions. The source is here imaged using the relative sound intensity measured at each direction in front of the array, while its corresponding real position in the grid is illustrated on the top side of the figure. From each of the intensity map, we can clearly identify the location information of the sound source. It is also important to point out that for this experiment the camera image was not calibrated with the array output and therefore, the results shown in Figure 19 are not absolute but relative. Nevertheless, it is possible to observe that even with this un-calibrated localization, important information about the sound intensity distribution can still be obtained by our microphone array system.

To calculate the position errors, we measured the coordinate of the peak value from each of the sound intensity map, and used this information as the estimated source location. By calculating the distances between the coordinates at the four corners, and mapping these distances onto an 8 by 8 grid with a spacing of 5 degrees (or 2.3 cm), we were able to convert the coordinates acquired from Figure 19 bottom to the same local coordinate system used by the real positions of the sound source. Compared to their real locations, the mean error of the estimated positions is 1.1 degrees (or 0.49 cm) with a maximum error of 3.1 degrees (or 1.44 cm). This demonstrates that our microphone array system has the capability of accurately estimating the location of the sound source.

5.4.2 Sound Source Separation

Another case study to demonstrate the array system's performance is illustrated by Figure 20. In this experiment, the sound sources were two omnidirectional speakers (3.2 cm in diameter) emitting a single tone of 5 kHz at the same time and located 40 cm away from the microphone array. The spacing between the two sound sources was set to be 6.3 cm in the beginning, and then increased by 2 cm after each recording. Similar to source localization, all 64 channels of audio signals were sampled at 50 kHz for 1 second, and then processed by the DS beamformer in order to measure the sound intensities at every direction in front of the array. The results can be seen in Figure 20. Notice that the camera was not aligned with the array during experiment, thus can only provide relative location information.

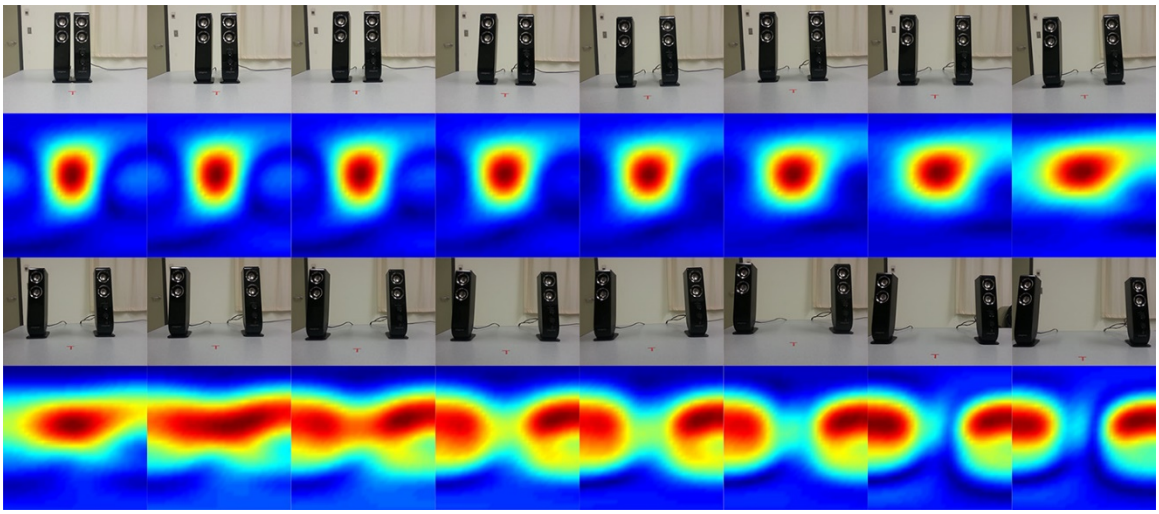

Figure 20 Separation of two source signals whose spacing ranges from 6.3 cm to 36.3 cm.

As shown in the obtained intensity maps, with the increase of the distance between the two speakers, the image of the sources becomes wider, and after a certain point

(approximately 24 cm in this case), it begins to separate into two, and eventually we are able to identify the presence of the two sound sources. Using similar approach described in source localization, a mean separation error of 13.1 degrees (or 1.31 cm) was obtained for this experiment. This was based on the comparison of the estimated and real coordinates after the two separated signals were observed (Figure 20 bottom). We will discuss these results, along with this "separation distance" in further details in the next chapter.

5.5 Imaging of Different Materials

To generate the acoustic images and test the frequency responses of different materials, we illuminated the object to be analyzed by a single sound source standing at a fixed position near the array. By this, we were able to image the scene by processing the back-scattered reflections from the object and analyze its acoustic response for a range of frequencies. In Figure 22, we can see the reconstructed images based on the DSBF algorithm.

For this experiment, we used the same setup of the sound localization (Figure 21). The distance between the array and the object was 30 cm, and all the testing materials were located within the 8×8 grid on the wooden board, which made an imaging area of 30 cm L × 20 cm W × 20 cm H. The speaker was located in line with the surface of the array, and 5 cm to its left edge, and facing roughly 28 degree towards its right side in order to transmit a signal that can be reflected from the imaging area. The sound source signal

used was a continuous sinusoidal wave of frequencies from 1 kHz to 7 kHz, generated by the arbitrary waveforms in MATLAB. The output power of the speaker was 200 mW.

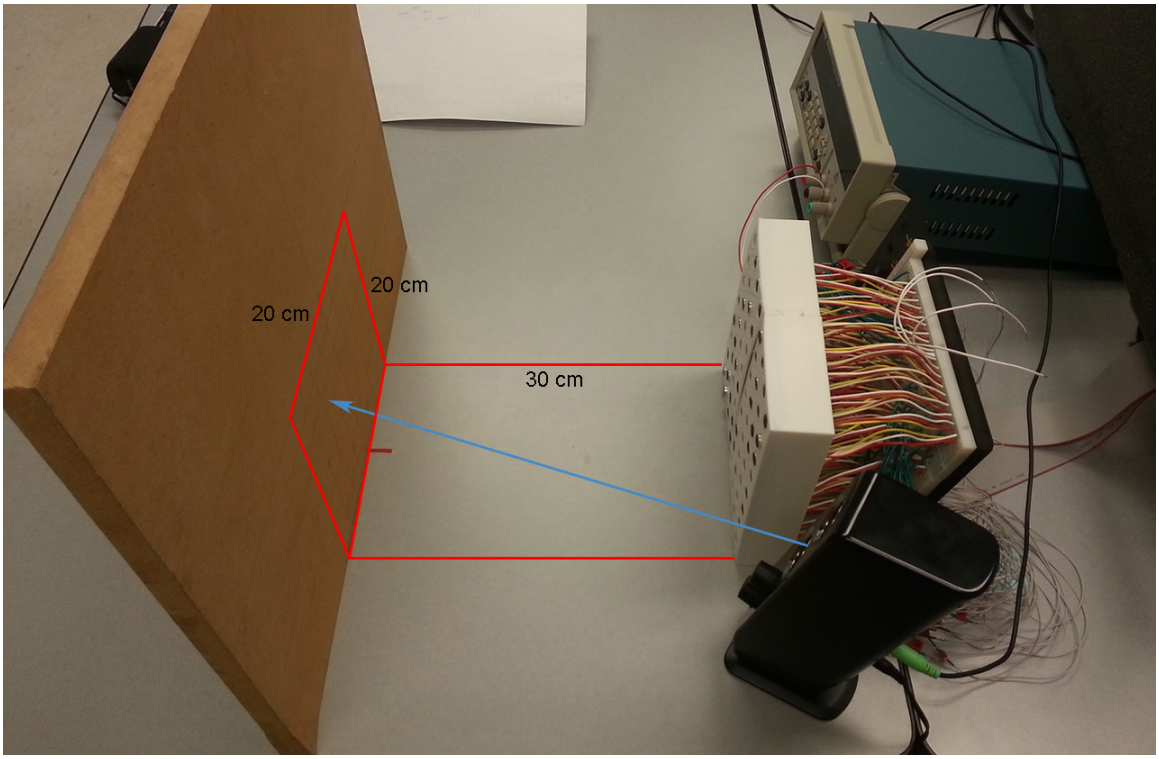

Figure 21 Setup for the imaging experiment. The speaker was located 5 cm to the left edge of the array, and the object was 30 cm in front of the array.

We chose wood, cardboard, metal, plastic, cloth, and rubber for testing materials as they were most commonly seen in our everyday lives. We also tested the imaging system with a human's hand in order to analyze its performance for objects with sophisticated textures. Each material was imaged using a range of frequencies from 1 kHz to 7 kHz, with a 0.5 kHz interval. As discussed in the previous chapters, to avoid spatial aliasing, the maximum frequency for the transmitted signal should not be higher than 7391.3 Hz for a sensor spacing of 2.3 cm. Thus, in our system, it was set to 7000 Hz. The reflected

signals from the tested object were received by the 64-node array, and then processed by the DSBF algorithm. The image was generated using the same procedure described in sound localization.

The imaged results of the 7 materials under different frequencies are shown in Figure 22. The same wooden board (48 cm × 41.5 cm) with 8×8 grid were used for both testing object and calibration. We first recorded the reflected signals from this wooden board, and obtained its sound images. By comparing these images with the results of the sound localization experiment (see Figure 19), we were able to relatively locate where the source signal were reflected from in the grid. All the other materials with smaller size were then glued over this specific area in the grid. This allowed us to calibrate the positions between the object and the sound source so that the signals received by the array were the direct reflections from the object itself.

From the results, we see that generally all testing materials have similar visual response patterns with the increase of the frequency. For frequencies above 3 kHz, we can clearly identify the location of the main peak which represents the direct reflection from the object. However, under low frequencies (1 kHz to 3 kHz), the main peak location is not fixed, and thus this value cannot be used as the reflection power from the object. Furthermore, we can observe that under some certain frequencies (such as 4 kHz, 5.5 kHz, and 6 kHz), the acoustic responses from human skin are relatively different compared to other materials. Additionally, the responses from cardboard at 4 kHz, 5 kHz, and 7 kHz are also visually distinctive. Although all the other materials may have similar

response patterns, we cannot compare their corresponding response powers here, as the scales of the color map for each image are not unified. Therefore, the same colors across different images do not represent the same value.

To further analyze the acoustic responses between different materials, we have plotted the frequency response curves of these materials in Figure 23, according to their relative signal intensities at the reflection area. As stated above, for frequencies above 3 kHz, the peak value in the image is the response power from the object, while for frequencies between 1 kHz to 3 kHz, we need to first locate the relative reflection area in its acoustic image by utilizing the object location information we acquired during the calibration process, and then measure the signal intensity level in the middle of this area as the estimation of the object's response power. As the plot shows, for each material, the reflection power varies significantly between different frequencies, and for each testing frequency, the textures between different objects will affect their response powers. With the increase of the frequency, there is a downward trend in general after 3500 Hz. While several materials have similar response curves, human skin and cardboard demonstrate a rather different responses, especially within some certain frequency ranges. Their acoustic responses are relatively weak from 3.5 kHz to 6.5 kHz. This is likely due to their rough textures, and we will discuss more about these results in Chapter 6.

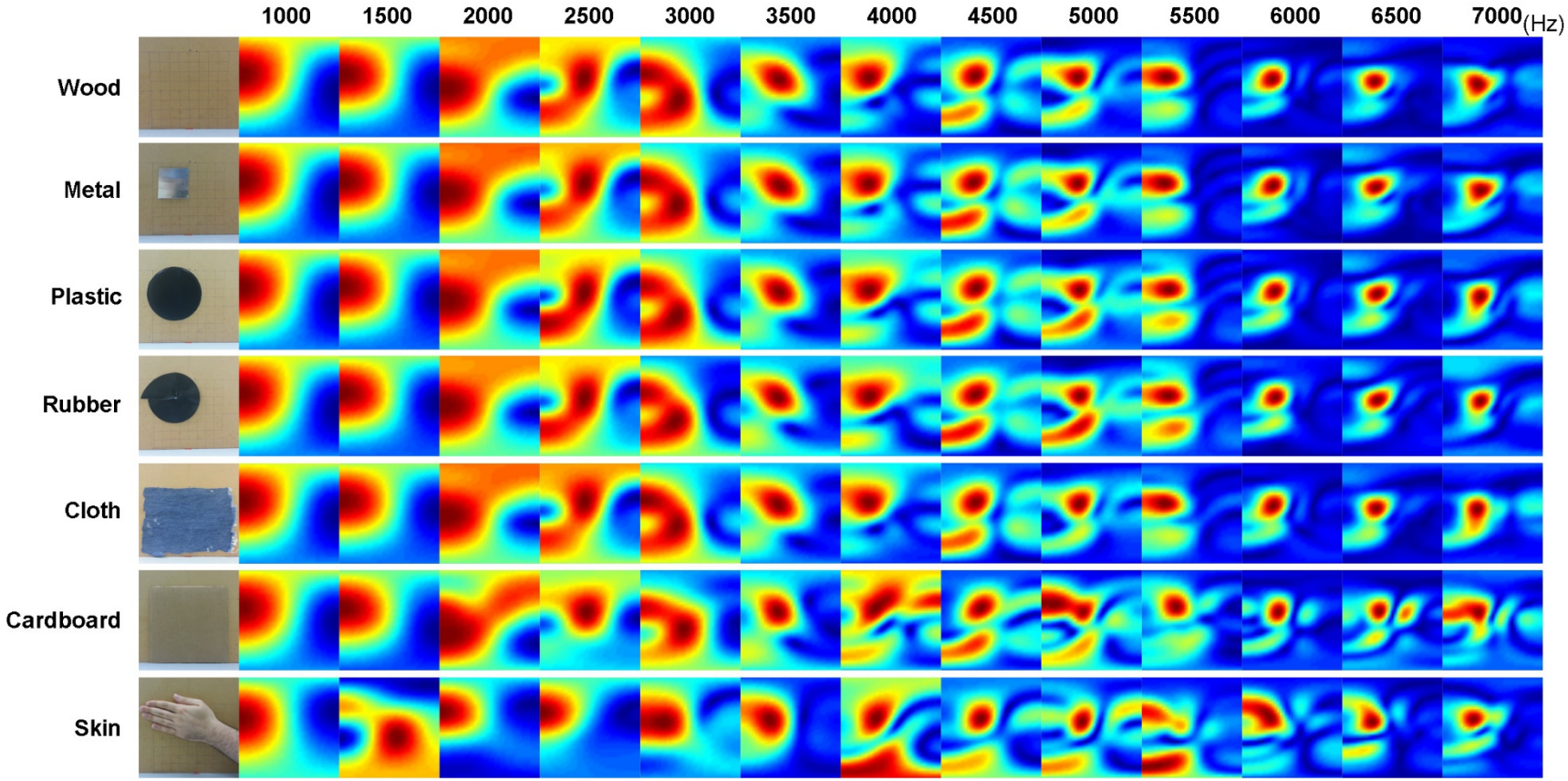

Figure 22 Reconstructed acoustic images of 7 different materials using a range of frequencies.

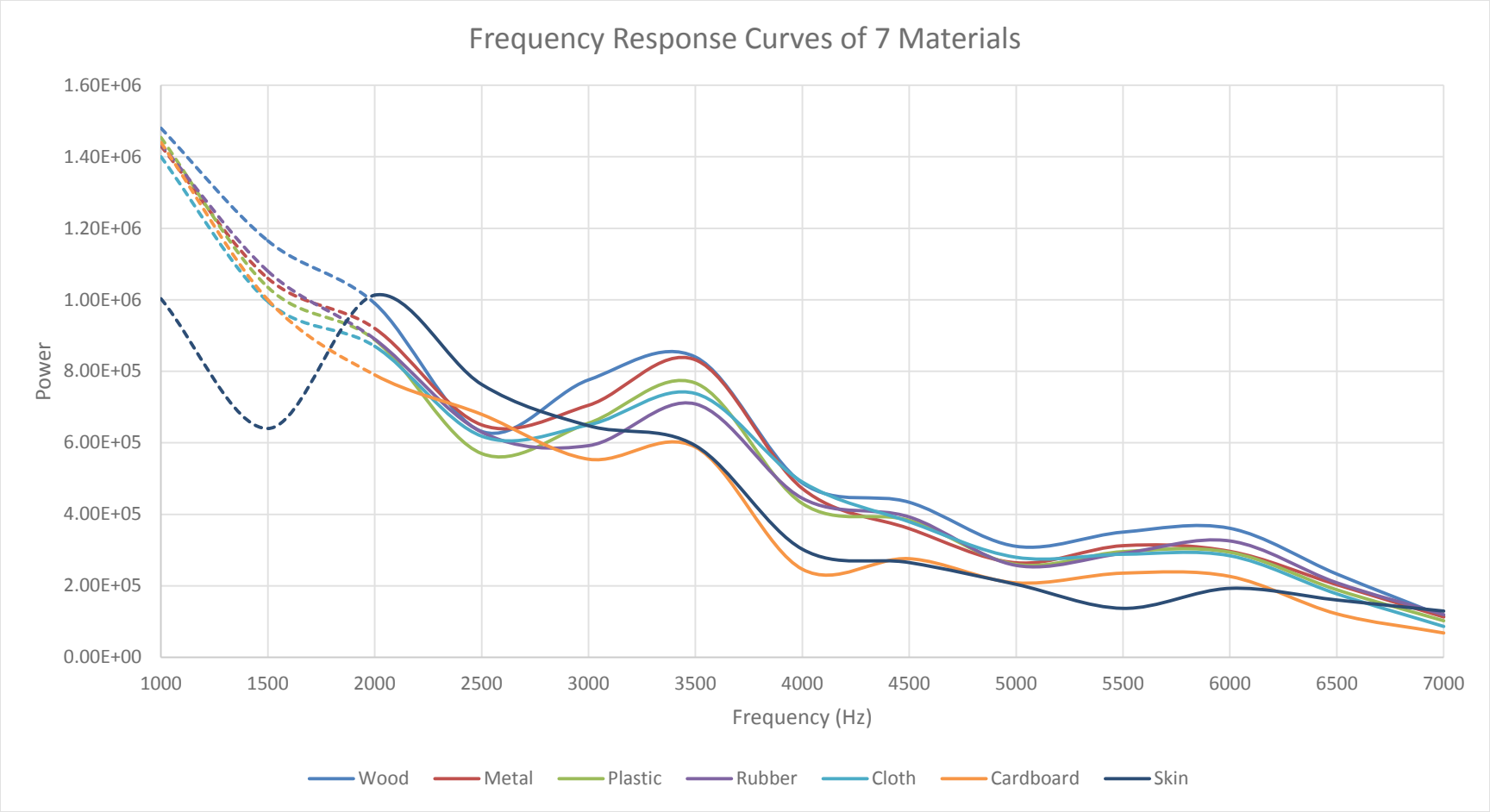

Figure 23 Frequency response curves of 7 different materials.

6 Chapter: Discussion

6.1 Performance of Source Localization and Separation

6.1.1 Localization Performance

From the results illustrated in Figure 19 and Figure 20, we demonstrate that our microphone array system is capable of estimating the location of sound sources and separating signals coming from different directions. Unlike traditional localization and separation techniques, which usually involve using a camera to track the sound sources in order to determine the look direction of the array, our system does not require any prior knowledge of the source locations, nor any particular information regarding to the source of noise. A mean position error of 1.1 degrees (or 0.49 cm) in the acoustic maps was obtained. This localization difference is less than $1/6^{th}$ the diameter of the speaker (3 cm) and less than $1/4^{th}$ the spacing of the acoustic map grid (5 degrees), which indicates the high accuracy of our microphone array system in localizing the sound source.

All the data was recorded in the presence of noise and reverberation, therefore it is likely that the reflected signals from other objects will also be captured and imaged by the array system. As can be seen in the first row of the localization results, except for the source image, the reflections from the table also appear. Although utilizing the delay-and-sum beamformer will attenuate this interference, it is still rather difficult to be eliminated, especially in everyday surroundings such as our testing environment. Nevertheless, this multipath issue does not affect the overall image quality of the sound source, which

shows the robustness and effectiveness of our system in sound localization. However, it is difficult to further determine how different noise conditions would affect the results of our experiments, since we did not setup different types of noise as comparison, and our noise sources were merely low-frequency environmental noise from computer cooling fans and an air conditioner. A formal experiment would be required in the future to verify such effect.

As stated in section 5.4.1, the camera images presented in the results were not calibrated with the generated intensity maps. This is due to the fact that the algorithms we used are data-driven, which means it does not require the prior location information from a camera. The results show that we can still obtain relatively accurate sound source positions. With the source coordinates located on the edge, we are able to measure the average angular detection range of our array at 30 cm, which is from -30 to 25 degree in azimuth, and 25 to -25 degree in elevation. This information is crucial to applications such as range detection and navigation. However, as the localization experiment was only conducted in a 2D plane, we did not acquire any information regarding to the depth of the sound source. This information can be measured based on the intensity level of the imaged source signal, since the signal intensity would generally decrease with the increase of its distance to the array. Future experiments would be needed to further examine their relationships.

6.1.2 Separation Performance

Similar to localization, sound source separation with our system does not need prior information on the locations of the source signals. Our results give an average error of 13.1 degrees (or 1.31 cm) for the separation accuracy, which is comparable to the increased distance between the two speakers for each recording (2 cm). This error was calculated after the two source signals became identifiable in the acoustic image, since the system could not separate the sound sources when their distance was too close.

As shown in Figure 20, we notice that the two source signals are mixed in the resulting image when the spacing between the speakers is less than 24 cm, and after this distance, the separated source images are becoming identifiable, allowing us to locate their relative positions. This "separation distance" is possibly related to the beamwidth of the mainlobe in the array's beam pattern. The source frequency used for this experiment was 5 kHz, and as discussed in Chapter 3.6, the mainlobe width is a function of the sensor number, sensor spacing, and signal frequency. When the number of sensors and the spacing of sensors are fixed, it will only decrease with the increase of the signal frequency, which means higher frequency will make a sharper mainlobe. To illustrate this, we plot the beam patterns of our array when the frequency is increased from 1 kHz to 7 kHz in Figure 24. When the spacing of the sound sources is less than (or comparable to) the beamwidth of the mainlobe, the beamformer cannot separate them from each other. To reduce the separation distance, we need a sharper mainlobe, and thus, a higher frequency. However, this will also introduce the problem of spatial aliasing. In the future we will

address this issue by incorporating blind source separation techniques such as independent component analysis [77].

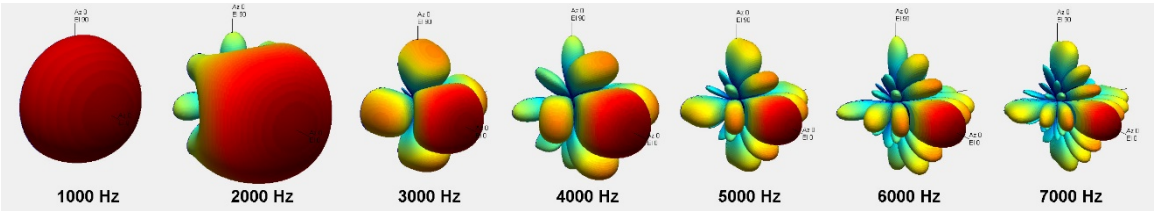

Figure 24 Beam patterns with the increase of the frequency from 1 kHz to 7 kHz. The look direction of the array is at 0 Az, 0 El.

6.1.3 Image Resolution

The resolution of the reconstructed images depends on two main factors. First, as discussed in the above section, the mainlobe width determines the accuracy of the source localization and separation. A narrower mainlobe will produce a sharper source image in the generated intensity map. Second, the reconstructed sound image is created by steering the beamformer along a set of 1369 directions (a 37×37 grid in azimuth and elevation), and quantizing the steered response power according to a color map. Therefore, by reducing the scan interval (currently it is 5 degrees between each direction), we are able to obtain a higher spatial resolution. However, this will also increase the system's overall computational power and time, which would become impractical for real-time applications.

6.2 Acoustic Response of Different Materials

Our experiments concentrated on testing the frequency response of typical objects as a basis for the navigation system of a mobile robot. For these tests, we chose seven targets: wood, cardboard, metal, plastic, cloth, rubber, and human hand. These targets have different textures, therefore, they were used to examine the system's ability to distinguish the target and assess its response.

The obtained acoustic images of all seven materials are shown in Figure 22. While the main peaks from frequencies above 3 kHz are clearly the direct reflection from the objects themselves, we can also see echoes from the bottom edges (such as those in 4.5 kHz, 5 kHz, and 5.5 kHz). These echoes do not belong to the objects, but are attenuated echoes from the wooden board, as they have relatively the same power intensities compared to the ones from the wood. We have calibrated the position of the object in order to guarantee the direct reflections are always on the analyzed target.

However, in the low frequency range (1 kHz to 3 kHz), the signal from the sound source overlaps with the direct reflection, and thus the main peak no longer represents the response power of the object. This is possibly due to the fact that the beamwidth of the mainlobe in low frequencies is relatively wide (see Figure 24), as a result, the array does not possess the precision sufficient to distinguish the reflected signals from its source. To solve this issue, we need to utilize the location information we acquired during the calibration process (see section 5.5), and find out the relative reflection areas in order to measure the signal intensity from the target. Another issue with low-frequency sound

source is that its wavelength is relatively long. For example, at 1 kHz, the wavelength is about 34.2 cm, which is larger than the size of the testing object. This will cause diffraction, which means the sound wave will no longer be reflected from the surface of the material. Even if we can measure the signal power based on the localized reflection area, this value does not necessarily represent the real response power from the object, which makes the results in low frequencies (especially from 1 kHz to 2 kHz) relatively inaccurate.

The acquired frequency response curves of all seven materials are shown in Figure 23. The curves between 1 kHz and 2 kHz are plotted in dot line because they may not be accurate. It can be seen that each material has its unique response with the increase of the frequency. Overall, wood, metal, plastic, cloth, and rubber have good sound reflecting properties, while the signal power from cardboard and human skin is relatively weak from 3.5 kHz to 6.5 kHz. This is likely due to the complex texture of the human skin (and cardboard), which scatters or absorbs most of the energy so that the detected echo represents only a small portion of the original signal. The result indicates that generally different textures will have different acoustic response powers. Our system can detect the difference between textures based on their unique response powers at certain frequency. More specifically, from 3.7 kHz to 4.3 kHz, and 5 kHz to 6 kHz, the response powers of human skin and cardboard present a significant difference compared to other materials. Such differences can also be observed in Figure 22 at 4 kHz, 5 kHz and 5.5 kHz, where the sound images generated from the human skin and cardboard can be clearly distinguished from other materials.

The object classes that can be present in the scene vary greatly and are difficult to model by only a certain amount of sample materials, because the representation of the direct reflection depends on both the relative position of the object to the array as well as the angle from which the excitation signal is emitted into the scene. The response of the reflected echo should depend on both the texture and shape of the reflecting surface and its orientation relative to the array. However, our experiments only examined the effect of the texture. As we see from Figure 22, all testing materials are not in the same size. While wood, cardboard, and cloth have larger size, metal, plastic, rubber, and human hand are relatively small. Despite the fact that the size of the object varies, we still acquired similar acoustic responses as the frequency is increased. A possible explanation would be that since we are using audible sound, its wavelength is still relatively long even at higher frequencies (such as 7 kHz), which makes it harder to be reflected from the edge of the material. Therefore we can only observe its peak reflection from the center of the object. Although there exist some distinct acoustic phenomena from rough surfaces (cardboard and human skin) at certain frequencies, we cannot exclude the factor of the object's shape, especially for the human hand, since it has a more complex 3D shape, and this may have more influence on its response pattern. Of course, more experiments would be needed in the future in order to evaluate the effect of the object's shape.

6.3 Summary

In this work, we focused on the design of the system, and did not implement sophisticated beamforming algorithms or other signal processing software components.

However, even with the simplest beamformer possible, we were able to obtain relatively accurate results of sound source localization and separation, as well as detect the differences between textures based on their acoustic responses presented in the frequency response curves and signal intensity maps.

Comparison with past work is difficult for several reasons. One reason is differences in experimental conditions. Our data were collected in a very noisy and reverberant environment, likely noisier than most of the recently-published results. For example, Legg and Bradley [39] obtained a mean position difference of 0.66 cm for sound source localization in an anechoic chamber, using a 72-element spherical microphone array combined with a digital camera, while our work achieved a mean position error of 1.1 degrees (0.49 cm) in a laboratory room. Another reason is due to the different array geometries presented in other works which will have a strong impact on the results. For instance, Pei et al. [6] reported an average localization error of 25 cm along the horizontal axis and 53 cm error along the vertical axis, using a small-scale linear microphone array. Our results have obtained a significant improvement with a 64-node rectangular array, which demonstrates the benefits of having a larger array size. Finally, most recently-published results of object detection experiments were based on ultrasound, which has a much higher frequency (and thus accuracy) compared to ours. For example, Moebus and Zoubir [11] compared the differences between reflections from a polyvinyl chloride (PVC) pole with a smooth surface and a rough surface covered with bubble wrap, using a 20-by-20 rectangular array and 50 kHz ultrasound. Their experiments showed a strong reflection from a small fraction of the smooth surface, and a weak reflection from the

whole object with rough surface. Similar to their results, our work can also detect the differences between smooth and rough textures, with audible sound and a smaller array.

6.4 Potential Application for Mobile Robot Navigation

Real world applications for mobile robot navigation demand very detailed sensor information to provide the robot with good environment-interaction capabilities [79]. Vision-based sensors can provide the robot with a significant amount of information about its surroundings, and thus are potentially the most powerful source of information among all other sensors used on robots. At present, high resolution optical sensors are most widely adopted for mobile robot positioning and navigation [84]. Such sensors are capable of capturing image features that match the landmarks or maps of the environment.

In order to navigate in an environment, a robot needs to use its sensors to create a map (or representation) of the local environment. This local map is then compared to the global map previously stored in memory. If a match is found then the robot can compute its actual position and orientation in the environment. The pre-stored map can be a computer-aided drafting (CAD) model of the environment, or it can be constructed from prior sensor data.

Within this context, our microphone array imaging system can prove useful for robot navigation, and further, object detection. For example, based on the unique response patterns between smooth and rough textures presented in the reconstructed acoustic maps

at certain frequencies, we can create a pre-stored map as the representation of typical object such as wooden door (smooth texture) or cardboard box (rough texture). If the array system detects a response that matches this map, the robot will interpret it as detecting this object, and will respond according to the programmed commands. Furthermore, with the frequency response curves acquired, we could quantify the response of different objects at certain frequency. This information can then be utilized to generate an updated map of the environment, which allows a robot to learn about a new environment and to improve positioning accuracy through exploration.

In comparison with optical and ultrasonic based navigation systems, our microphone array system has the potential to provide a simple and cheap sensor alternative for mobile robot navigation, which may allow for relatively precise localization as well as object detection. However, due to the scope of this thesis and time limitation, we did not actually implement our microphone array system in a mobile robot in order to evaluate its performance for real navigation and object detection tasks. This capability will be examined in the future, with more experiments.

7 Chapter: Conclusions

7.1 Summary of Findings

In this thesis, we have examined the design, implementation, and evaluation of a 64-node microphone array system that is capable of achieving acoustic imaging. We have outlined the design procedure of the array hardware and software architecture, and our utilization of the delay-and-sum beamforming algorithm for processing the data and generating the acoustic image. In addition, we have evaluated the performance of our array system in sound source localization and separation, and also tested its ability to image and detect seven different materials using audible sound.

Our microphone array system is able to locate the sound source with an average error of 1.1 degrees. It can also separate sound sources when their spacing is larger than the mainlobe width of the array, with a mean error of 13.1 degrees. Both experiments were conducted in the presence of noise and reverberation, which proves our system's robustness for real world applications.

We have presented the acoustic images of seven different materials as well as their frequency response curves from 1 kHz to 7 kHz. The reconstructed images generally show a similar visual response pattern with the increase of frequency, while a relatively distinct response can be observed from human skin and cardboard at certain frequency. Furthermore, from the frequency response curves, we are able to acquire the reflection

power of different materials from 2 kHz to 7 kHz, which can then be utilized to distinguish the objects. Our results show that generally materials with smooth texture have a strong response power, and those with rough texture have a relatively weak response. These unique responses will help to build the representation of different objects in the environment, which can be useful for mobile robot navigation such as object detection and recognition.

Through these experiments, we have demonstrated that our microphone array system has the capability of reconstructing the acoustic scene with audible sound in order to provide the location information of the sources as well as the acoustic responses of different targets. We can conclude that our system is a cheap, low-power alternative technology for acoustic imaging.

7.2 Limitations and Future Work

Currently, the delay-and-sum beamforming algorithm has to be operated off-line due to the expensive cost in the computation of the spatial power spectrum of the acoustic scene, which was computed by steering the array at many directions, and performing the beamforming for each direction. Due to current host interface constraints, it was not practical to record the output of all 64 microphones and process the data in real-time. However, this capability can be achieved in the future, using custom-built hardware architecture such as the FPGA platform.

In addition, in order to evaluate the quantitative performance of the array further, we will combine our system with depth-sensing cameras so that we can map the obtained acoustic images onto a 3D scene of the environment. This will allow us to calculate the positioning errors more accurately, as well as compare the acoustic response of the object with its real-world entity in more detail.

Another important aspect that we plan to consider in the future is integrating our imaging array system with map matching algorithms in order to examine its efficacy in utilizing the acoustic responses from different textures to detect the presence of their corresponding objects. We also plan to collect more reflection patterns from more different textures with various sizes so that we can create a diversified database for such map-based navigation system. Additionally, since the sound source itself (such as its power, its distance and orientation relative to the array) and also the shape of the object may affect the reflections, we hope to explore these factors and understand their effects on the acoustic responses.

Finally, while we have shown that the delay-and-sum beamforming algorithm presented in this thesis is capable of generating acoustic images in noisy and reverberant environments, its performance in other applications has yet to be fully explored. Therefore, we will consider applying our microphone array system to other domains such as speech enhancement and gesture sensing.

References

[1]. Korpel, A. (1968). Acoustic imaging and holography. *Spectrum, IEEE, 5*(10), 45-52.

[2]. Aarabi, P. (2003). The fusion of distributed microphone arrays for sound localization. *EURASIP Journal on Applied Signal Processing, 2003,* 338-347.

[3]. Schmidt, R. O. (1986). Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on, 34*(3), 276-280.

[4]. Goseki, M., Takemura, H., & Mizoguchi, H. (2011, December). Visualizing sound pressure distribution by kinect and microphone array. In *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on* (pp. 1243-1248). IEEE.

[5]. Goseki, M., Ding, M., Takemura, H., & Mizoguchi, H. (2011, October). Combination of microphone array processing and camera image processing for visualizing sound pressure distribution. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on* (pp. 139-143). IEEE.

[6]. Pei, L., Chen, L., Guinness, R., Liu, J., Kuusniemi, H., Chen, Y., ... & Soderholm, S. (2013, October). Sound positioning using a small-scale linear microphone array. In *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on* (pp. 1-7). IEEE.

[7]. Turqueti, M., Saniie, J., & Oruklu, E. (2010, August). MEMS acoustic array embedded in an FPGA based data acquisition and signal processing system. In *Circuits and Systems (MWSCAS), 2010 53rd IEEE International Midwest Symposium on* (pp. 1161-1164). IEEE.

[8]. Miyake, R., Hayashida, K., Nakayama, M., & Nishiura, T. (2013, June). A study on acoustic imaging based on beamformer to range spectra in the phase interference method. In *Proceedings of Meetings on Acoustics* (Vol. 19, No. 1, p. 055041). Acoustical Society of America.

[9]. Strakowski, M. R., Kosmowski, B. B., Kowalik, R., & Wierzba, P. (2006). An ultrasonic obstacle detector based on phase beamforming principles. *Sensors Journal, IEEE, 6*(1), 179-186.

[10]. Harput, S., & Bozkurt, A. (2008). Ultrasonic phased array device for acoustic imaging in air. *Sensors Journal, IEEE, 8*(11), 1755-1762.

[11]. Moebus, M., & Zoubir, A. M. (2007, April). Three-dimensional ultrasound imaging in air using a 2D array on a fixed platform. In *Acoustics, Speech and Signal*

Processing, 2007. ICASSP 2007. IEEE International Conference on (Vol. 2, pp. II-961). IEEE.

[12]. Moebus, M., Zoubir, A. M., & Viberg, M. (2010, March). Parametrization of acoustic images for the detection of human presence by mobile platforms. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* (pp. 3538-3541). IEEE.

[13]. Dokmanic, I., & Tashev, I. (2014, May). Hardware and algorithms for ultrasonic depth imaging. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 6702-6706). IEEE.

[14]. Ultrasonic Audio Technologies. (2015). Acouspade [Online]. Retrieved from http://www.ultrasonic-audio.com/products/acouspade.html.

[15]. Nakayama, Y., Adachi, M., Ishimoto, K., & Tatekura, Y. (2013, July). Individual sound image generation for multiple users based on loudspeaker array with NBSFC. In *Digital Signal Processing (DSP), 2013 18th International Conference on* (pp. 1-5). IEEE.

[16]. Liang, B., Yuan, B., & Cheng, J. C. (2009). Acoustic diode: Rectification of acoustic energy flux in one-dimensional systems. *Physical review letters, 103*(10), 104301.

[17]. Liang, B., Guo, X. S., Tu, J., Zhang, D., & Cheng, J. C. (2010). An acoustic rectifier. *Nature materials, 9*(12), 989-992.

[18]. Thiergart, O., Del Galdo, G., Taseska, M., & Habets, E. (2013). Geometry-based spatial sound acquisition using distributed microphone arrays. *Audio, Speech, and Language Processing, IEEE Transactions on, 21*(12), 2583-2594.

[19]. Hoflinger, F., Zhang, R., Hoppe, J., Bannoura, A., Reindl, L. M., Wendeberg, J., ... & Schindelhauer, C. (2012, November). Acoustic Self-calibrating System for Indoor Smartphone Tracking (ASSIST). *In Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on* (pp. 1-9). IEEE.

[20]. Bischof, G. (2008, October). Acoustic imaging of sound sources-a junior year student research project. In *Frontiers in Education Conference, 2008. FIE 2008. 38th Annual* (pp. S2C-1). IEEE.

[21]. Qin, K., & Li, Y. (2014, May). Real-time ultrasonic imaging for multi-layered objects with synthetic aperture focusing technique. In *Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, 2014 IEEE International* (pp. 561-566). IEEE.

[22]. Stride, E., & Saffari, N. (2003). Microbubble ultrasound contrast agents: a review. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 217*(6), 429-447.

[23]. Bruce, M., Averkiou, M., & Powers, J. (2007). Ultrasound contrast in general imaging research. *Eindhoven: Philips Medical Systems,* 1-20.

[24]. Maeda, Y., Sugimoto, M., & Hashizume, H. (2011, October). A robust doppler ultrasonic 3D imaging system with MEMS microphone array and configurable processor. In *Ultrasonics Symposium (IUS), 2011 IEEE International* (pp. 1968-1971). IEEE.

[25]. Van Veen, B. D., & Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. *IEEE assp magazine, 5*(2), 4-24.

[26]. Wall, K., & Lockwood, G. R. (2005, September). Modern implementation of a realtime 3D beamformer and scan converter system. In *Ultrasonics Symposium, 2005 IEEE* (Vol. 2, pp. 1400-1403). IEEE.

[27]. Microsoft. (2014). Kinect for Windows [Online]. Retrieved from http://www.microsoft.com/en-us/kinectforwindows/.

[28]. Nokia. (2013). Lumia 925 [Online]. Retrieved from http://www.microsoft.com/en/mobile/phone/lumia925/.

[29]. Acoustic Camera. (2014). Ring 72-120 AC Pro [Online]. Retrieved from http://www.acoustic-camera.com/.

[30]. Polycom. (2013). CX5100 Unified Conference Station [Online]. Retrieved from http://www.polycom.com/products-services/products-for-microsoft/lync-optimized/cx5100-unified-conference-station.html.

[31]. Weinstein, E., Steele, K., Agarwal, A., & Glass, J. (2004). LOUD: A 1020-Node Modular Microphone Array and Beamformer for Intelligent Computing Spaces.

[32]. MIT CSAIL. (2004). MIT Project Oxygen [Online]. Retrieved from http://oxygen.lcs.mit.edu/.

[33]. Theodoropoulos, D., Kuzmanov, G., & Gaydadjiev, G. (2013). Custom architecture for multicore audio beamforming systems. *ACM Transactions on Embedded Computing Systems (TECS), 13*(2), 19.

[34]. Donovan, A. O., Duraiswami, R., & Zotkin, D. (2008, March). Imaging concert hall acoustics using visual and audio cameras. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on* (pp. 5284-5287). IEEE.

[35]. Gupta, S., Morris, D., Patel, S., & Tan, D. (2012, May). Soundwave: using the doppler effect to sense gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1911-1914). ACM.

[36]. Bedri, H., Feigin, M., Everett, M., Charvat, G. L., & Raskar, R. (2014, July). Seeing around corners with a mobile phone?: synthetic aperture audio imaging. In *ACM SIGGRAPH 2014 Posters* (p. 84). ACM.

[37]. Meyer, A., Döbler, D., Hambrecht, J., & Matern, M. (2011, June). Acoustic Mapping on three-dimensional models. In *Proceedings of the 12th International Conference on Computer Systems and Technologie*s (pp. 216-220). ACM.

[38]. Jing, Z., Bo, L., LU, D., & Errui, C. (2011, July). An acoustic imaging simulation based on microphone array. In *Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC), 2011* (Vol. 2, pp. 1398-1401). IEEE.

[39]. Legg, M., & Bradley, S. (2013). A combined microphone and camera calibration technique with application to acoustic imaging. *Image Processing, IEEE Transactions on, 22*(10), 4028-4039.

[40]. Kalgaonkar, K., & Raj, B. (2009, April). One-handed gesture recognition using ultrasonic Doppler sonar. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 1889-1892). IEEE.

[41]. Huang, X. (2009). Real-time algorithm for acoustic imaging with a microphone array. *The Journal of the Acoustical Society of America, 125*(5), EL190-EL195.

[42]. Wehr, T. M., & Wehr, J. R. (2014). Focusing sound waves using a two-dimensional nonlinear system. *J Emerg Inv.*

[43]. Jiang, D., Yi, J., & Bian, G. (2011, June). A new method of target focused sound. In *Computer Science and Service System (CSSS), 2011 International Conference on* (pp. 1957-1959). IEEE.

[44]. Candelas, P., Rubio, C., Gómez-Lozano, V., Belmar, F., & Uris, A. (2014). Acoustic lens based on a subwavelength slit surrounded with grooves. *Forum Acustisum 2014*.

[45]. Zhu, J., Christensen, J., Jung, J., Martin-Moreno, L., Yin, X., Fok, L., ... & Garcia-Vidal, F. J. (2011). A holey-structured metamaterial for acoustic deep-subwavelength imaging. *Nature physics, 7*(1), 52-55.

[46]. Sánchez-Dehesa, J., Caballero, D., Cervera, F., Sanchis, L., Sánchez-Perez, J. V., Martinez-Sala, R., ... & Lopez, C. (2002). Refractive acoustic devices for airborne sound. *The Journal of the Acoustical Society of America, 112*(5), 2413-2413.

[47]. He, Z., Deng, K., Zhao, H., & Li, X. (2012). Designable hybrid sonic crystals for transportation and division of acoustic images. *Applied Physics Letters, 101*(24), 243510.

[48]. Sánchez-Pérez, J. V., Caballero, D., Martinez-Sala, R., Rubio, C., Sánchez-Dehesa, J., Meseguer, F., ... & Gálvez, F. (1998). Sound attenuation by a two-dimensional array of rigid cylinders. *Physical Review Letters, 80*(24), 5325.

[49]. Miyashita, T. (2005). Sonic crystals and sonic wave-guides. *Measurement Science and Technology, 16*(5), R47.

[50]. Miyashita, T., Taniguchi, R., & Sakamoto, H. (2003). Experimental full band-gap of a sonic-crystal slab made of a 2D lattice of aluminum rods in air. *WCU, 9*(2003), 911-914.

[51]. Miyashita, T., & Inoue, C. (2001). Sonic-crystal wave-guides by acrylic cylinders in air experimental observations based on numerical analyses. In *Ultrasonics Symposium, 2001 IEEE* (Vol. 1, pp. 615-618). IEEE.

[52]. Ishiguro, Y., & Poupyrev, I. (2014, April). 3D printed interactive speakers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1733-1742). ACM.

[53]. Ishiguro, Y., Brockmeyer, E., Rothera, A., & Israr, A. (2014, October). Ubisonus: spatial freeform interactive speakers. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology* (pp. 67-68). ACM.

[54]. Chan, V., & Perlas, A. (2011). Basics of ultrasound imaging. In *Atlas of Ultrasound-Guided Procedures in Interventional Pain Management* (pp. 13-19). Springer New York.

[55]. Teutsch, H. (2007). *Modal array signal processing: principles and applications of acoustic wavefield decomposition* (Vol. 348). Springer.

[56]. Keating, P. N., Sawatari, T., & Zilinskas, G. (1979). Signal processing in acoustic imaging. *Proceedings of the IEEE, 67*(4), 496-510.

[57]. Benesty, J., Chen, J., & Huang, Y. (2008). *Microphone array signal processing* (Vol. 1). Springer Science & Business Media.

[58]. Turqueti, M., Oruklu, E., & Saniie, J. (2012). Smart acoustic sensor array (SASA) system for real-time sound processing applications. Retrieved from http://www.ece.iit.edu/~eoruklu/IIT/Publications_files/Review%20Copy%202.pdf.

[59]. Nehorai, A., & Paldi, E. (1994). Acoustic vector-sensor array processing. *Signal Processing, IEEE Transactions on, 42*(9), 2481-2491.

[60]. Johnson, D. H., & Dudgeon, D. E. (1993). *Array signal processing.* New Jersey: Prentice Hall.

[61]. Frost III, O. L. (1972). An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE, 60*(8), 926-935.

[62]. Griffiths, L. J., & Jim, C. W. (1982). An alternative approach to linearly constrained adaptive beamforming. *Antennas and Propagation, IEEE Transactions on, 30*(1), 27-34.

[63]. Seltzer, M. L. (2003). *Microphone array processing for robust speech recognition.* Doctoral dissertation, Carnegie Mellon University Pittsburgh, PA.

[64]. Kajbaf, H., & Ghassemian, H. (2009). Acoustic Imaging of Heart Using Microphone Arrays. In *13th International Conference on Biomedical Engineering* (pp. 738-741). Springer Berlin Heidelberg.

[65]. Adib, F., Kabelac, Z., Katabi, D., & Miller, R. C. (2014, April). 3d tracking via body radio reflections. In *Usenix NSDI* (Vol. 14).

[66]. Adib, F., Kabelac, Z., & Katabi, D. (2014). Multi-Person Motion Tracking via RF Body Reflections.

[67]. General Standards Corporation. (2011). PMC66-16AI64SSA/C analog input board [Online]. Retrieved from http://www.generalstandards.com/view-products2.php?BD_family=16ai64ssc.

[68]. MathWorks. (2015). MATLAB [Online]. Retrieved from http://www.mathworks.com/products/matlab/.

[69]. Silverman, H. F., Patterson, W. R., & Flanagan, J. L. (1998). The huge microphone array. *Concurrency, IEEE, 6*(4), 36-46.

[70]. Flanagan, J. L., Johnston, J. D., Zahn, R., & Elko, G. W. (1985). Computer-steered microphone arrays for sound transduction in large rooms. *The Journal of the Acoustical Society of America, 78*(5), 1508-1518.

[71]. Flanagan, J. L., Berkley, D. A., Elko, G. W., West, J. E., & Sondhi, M. M. (1991). Autodirective microphone systems. *Acta Acustica united with Acustica, 73*(2), 58-71.

[72]. Cutler, R., Rui, Y., Gupta, A., Cadiz, J. J., Tashev, I., He, L. W., ... & Silverberg, S. (2002, December). Distributed meetings: A meeting capture and broadcasting

system. In *Proceedings of the tenth ACM international conference on Multimedia* (pp. 503-512). ACM.

[73]. Sun, D., & Canny, J. (2012, September). A high accuracy, low-latency, scalable microphone-array system for conversation analysis. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 290-300). ACM.

[74]. Lang, S., Kleinehagenbrock, M., Hohenner, S., Fritsch, J., Fink, G. A., & Sagerer, G. (2003, November). Providing the basis for human-robot-interaction: A multi-modal attention system for a mobile robot. In *Proceedings of the 5th international conference on Multimodal interfaces* (pp. 28-35). ACM.

[75]. Taylor, M. B., Kim, J., Miller, J., Wentzlaff, D., Ghodrat, F., Greenwald, B., ... & Agarwal, A. (2002). The Raw microprocessor: A computational fabric for software circuits and general-purpose programs. *Micro, IEEE, 22*(2), 25-35.

[76]. Weinstein, E., Steele, K., Agarwal, A., & Glass, J. (2007, July). LOUD: A 1020-node microphone array and acoustic beamformer. In *International congress on sound and vibration (ICSV).*

[77]. Hyvärinen, A., Karhunen, J., & Oja, E. (2004). *Independent component analysis* (Vol. 46). John Wiley & Sons.

[78]. Dudgeon, D. E. (1977). Fundamentals of digital array processing. *Proceedings of the IEEE, 65*(6), 898-904.

[79]. Dixon, J., & Henlich, O. (1997). Mobile robot navigation. *Information Systems Engineering Year, Imperial College, 2,* 1-10.

[80]. Bardsley, B. G., & Christensen, D. A. (1981). Beam patterns from pulsed ultrasonic transducers using linear systems theory. *The Journal of the Acoustical Society of America, 69*(1), 25-30.

[81]. Brandstein, M., & Ward, D. (2001). *Microphone arrays: signal processing techniques and applications.* Springer Science & Business Media.

[82]. Johnson, D. H., & Dudgeon, D. E. (1992). *Array signal processing: concepts and techniques.* Simon & Schuster.

[83]. Adafruit. (2014). Electret microphone amplifier - MAX4466 with adjustable gain [Online]. Retrieved from https://www.adafruit.com/products/1063.

[84]. Fehlman, W. L., & Hinders, M. K. (2014). *Mobile robot navigation with intelligent infrared image interpretation.* Springer Science & Business Media.