Head Mounted Microphone Arrays

PHILIP W. GILLETT

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mechanical Engineering

Marty Johnson, Co-Chair
James Carneal, Co-Chair
Daniel Inman
Michael Roan
Craig Woolsey

August 27, 2009
Blacksburg, Virginia

KEYWORDS:
Acoustic Localization, Array Performance, Array Design Optimization, Voice Isolation,
Transparent Hearing

Head Mounted Microphone Arrays

Philip W. Gillett

(ABSTRACT)

Microphone arrays are becoming increasingly integrated into every facet of life. From sonar to gunshot detection systems to hearing aids, the performance of each system is enhanced when multi-sensor processing is implemented in lieu of single sensor processing. Head mounted microphone arrays have a broad spectrum of uses that follow the rigorous demands of human hearing. From noise cancellation to focused listening, from localization to classification of sound sources, any and all attributes of human hearing may be augmented through the use of microphone arrays and signal processing algorithms. Placing a set of headphones on a human provides several desirable features such as hearing protection, control over the acoustic environment (via headphone speakers), and a means of communication. The shortcoming of headphones is the complete occlusion of the pinnae (the ears), disrupting auditory cues utilized by humans for sound localization.

This thesis presents the underlying theory in designing microphone arrays placed on diffracting bodies, specifically the human head. A progression from simple to complex geometries chronicles the effect of diffracting structures on array manifold matrices. Experimental results validate theoretical and computational models showing that arrays mounted on diffracting structures provide better beamforming and localization performance than arrays mounted in the free field. Data independent, statistically optimal,

and adaptive beamforming methods are presented to cover a broad range of goals present in array applications. A framework is developed to determine the performance potential of microphone array designs regardless of geometric complexity. Directivity index, white noise gain, and singular value decomposition are all utilized as performance metrics for array comparisons. The biological basis for human hearing is presented as a fundamental attribute of headset array optimization methods. A method for optimizing microphone locations for the purpose of the recreation of HRTFs is presented, allowing transparent hearing (also called natural hearing restoration) to be performed. Results of psychoacoustic testing with a prototype headset array are presented and examined. Subjective testing shows statistically significant improvements over occluded localization when equipped with this new transparent hearing system prototype.

ACKNOWLEDGMENTS

I would like to thank Marty Johnson for his advice, patience, and guidance throughout the course of my work in the Vibration and Acoustics laboratories. His focus on delivering "the big picture" in every presentation and framing ideas thoroughly and completely in each paper have and will continue to improve my communication in the spoken and written word. Perhaps most of all I hope to have many more discussions about poverty, education, health care, politics, religion, economics, etc. in the future. Those previous exchanges of ideas, concepts, and predictions outside the field of engineering are some of my most enjoyable memories. For the record I still think Adam Smith had it right, but we can discuss that later.

Thanks also to Jamie Carneal for helping to provide challenging and interesting research applications. The times spent at Kentland Farms, VIR, and the field next to Plantation Road helped me see where and how my work could be applied. Great thanks to Jamie for giving me the opportunity to work for Brush Mountain Technologies to gain valuable experience in the many aspects of small business operations. Thanks also to Marten de Vries for his many insights and constant focus on commercialization, marketability, and the business aspect of each decision.

I would also like to thank Dr. Dan Inman, Dr. Craig Woolsey, and Dr. Michael Roan for all serving on my advisory committee, as well as for the classes they taught and the experiments they played a role in executing. I hope my work reflects the influence each of them had in my education, development, and research.

The work behind this research could not have been performed without funding from the Office of Naval Research, AuSIM Inc., Brush Mountain Technologies Inc., and Sennheiser

GmbH. Specifically I would like to thank Veronique Larcher and Dan Harris for the opportunity to continue improving designs and algorithms on (sometimes) very short notice.

Collaboration with other researchers was integral to the many experiments and applications carried on in my work: The lab certainly would be lost without a resident electrical engineer, Cory Papenfuss. There were countless times I am sure I interrupted his work to have electrical questions answered and he was always accommodating and helpful (most of the time it involved putting a capacitor in a circuit and having the circuit miraculously start working).

I do not think I could imagine a better match of temperaments and work ethics than I saw with Alessandro Toso and Thomas Funke. I am proud to have worked side by side with them in the lab and enjoyed our daily conversations on any number of topics. As if seeing Thomas's joy at talking about driving his Miata around were not enough, I was able to see that same joy when he sold it to Alessandro. I hope for nothing but the best for both of them in their future endeavors and hope that we will get to work together in the future.

I thank Dan Mennitt for the many discussions of experimental results and the efforts that produced those results. The fellow students who treated the main VAL lab as their office made for some of the most interesting conversations and enjoyable times: Brent Gold, Elizabeth Hoppe, Caroline Hutchison, Mike Milo, and Ryan Harne. Exchange students provided knowledge about European culture and a few German, French, and Italian phrases to go with it: Daniel Theurich and Sarah Gasser. Many of the German students to pass through the lab found their way with some help from the VAL cruise director: Kamal Idrisi. The VT Salsa club will never be the same.

Although I assured Dr. Chris Fuller that I would be retired by the time I was 27, I must amend that statement to occur at age 30, at the earliest. Talks with Dr. Fuller certainly helped lighten the mood as we exchanged experiences about some of the more curious aspects of life and politics.

The lab could not function with Gail Coe. The speed and efficiency with which Gail handled funding, contracts, and administrative concerns made the VAL work as it should: on research.

I would not be the person I am without my parents: Glenn and Susie Gillett. I thank them for their years of love, compassion, discipline, and direction that formed who I am today. My sisters Kate and Laura helped ground me in every situation, letting me know that however good or bad things went, I was still their brother. Many thanks to my aunts and uncles whose advice provided guidance they may not have even realized. To my Grandmommy and Granddaddy Gillett, I owe the focus on education and a life of learning. To Mema and my parents I owe the many experiences in which Faith and Hope were exemplified; the Faith and Hope that I then made my own. I thank my in-laws, Bob and Carol Brotherton for the unconditional love and support they have given me since the first day I met them. Lastly, and certainly most importantly, I thank the love of my life, Stacy, for the love, patience, support, and compassion she showed each and every day of my graduate education.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

CHAPTER 1

Introduction

1.1 Motivation

Microphone arrays are becoming increasingly integrated into every facet of life. From sonar to gunshot detection systems to hearing aids, the performance of each system is enhanced when multi-sensor processing is implemented in lieu of single sensor processing. Systems mounted on structures and vehicles generally have their niche applications, but human mounted, specifically head mounted, microphone arrays have a broad spectrum of uses aimed at augmenting or enhancing the abilities of human hearing. Applications potentially include noise cancellation, enhanced or focused listening, source localization, classification of sound sources, voice capture, and transparent hearing. All of these applications require specialized microphone array design and signal processing algorithms. The design of a head mounted array must take into account physical constraints of the mounting platform, the performance goals of the array, the effects on the wearer, as well as numerous weight, power, and size considerations. This thesis will focus on a couple of representative head mounted array geometries: a headset and a helmet.

1.1.1 Helmet Mounted

For military applications, soldiers would greatly benefit from accurate localization information about acoustic events. Unfortunately a very necessary safety structure interferes with the natural hearing capabilities of humans: the military helmet. In some

respects the additional acoustic diffraction caused by wearing a helmet is in many ways worse than that caused by wearing a set of headphones (discussed later). Rather than simply losing the ability to localize sound correctly, the helmet adversely affects directional cues to the point that one may still think he/she knows where the sound is coming from while being completely incorrect.

1.1.1.1 Example Application

The systems presented here are meant to provide an illustration of the end technology driving the research found in this dissertation. The first, and perhaps most obvious system, is an acoustic sensor array embedded within a soldier's helmet paired with appropriate data acquisition, signal processing, user interface, and power management systems. Figure 1.1 shows an initial prototype of such a system developed by Brush Mountain Technologies [1] (BMT) that uses separate systems for the filtering and acquisition of acoustic signals, a commercial, off-the-shelf (COTS) laptop for processing and communication, and a portable digital assistant (PDA) for displaying information. Other helmet mounted microphone arrays have been demonstrated by Bolt Beranek and Newman (BBN) [2], Vanderbilt University [3][4], and the US Army [5] for counter-sniper applications utilizing multiple arrays. These designs were all based on simple time delay of arrival (TDOA) methods that did not take advantage of the acoustical scattering effects caused by the helmet or human body. No mention of array design or microphone position optimization is mentioned and these array geometries appeared to simply serve the function of equipping a helmet with a microphone array. In one case the presence of a helmet is actually derided as causing undesired effects [5],

> "The main difference between the helmet array and open-structure array is that the helmet has sensors that are flushmounted to a masking surface, and sound must wrap around the helmet, producing multipath and shading."

The goals of this helmet mounted microphone array and processing system are to provide better localization accuracy and acoustic classification ability than accomplished

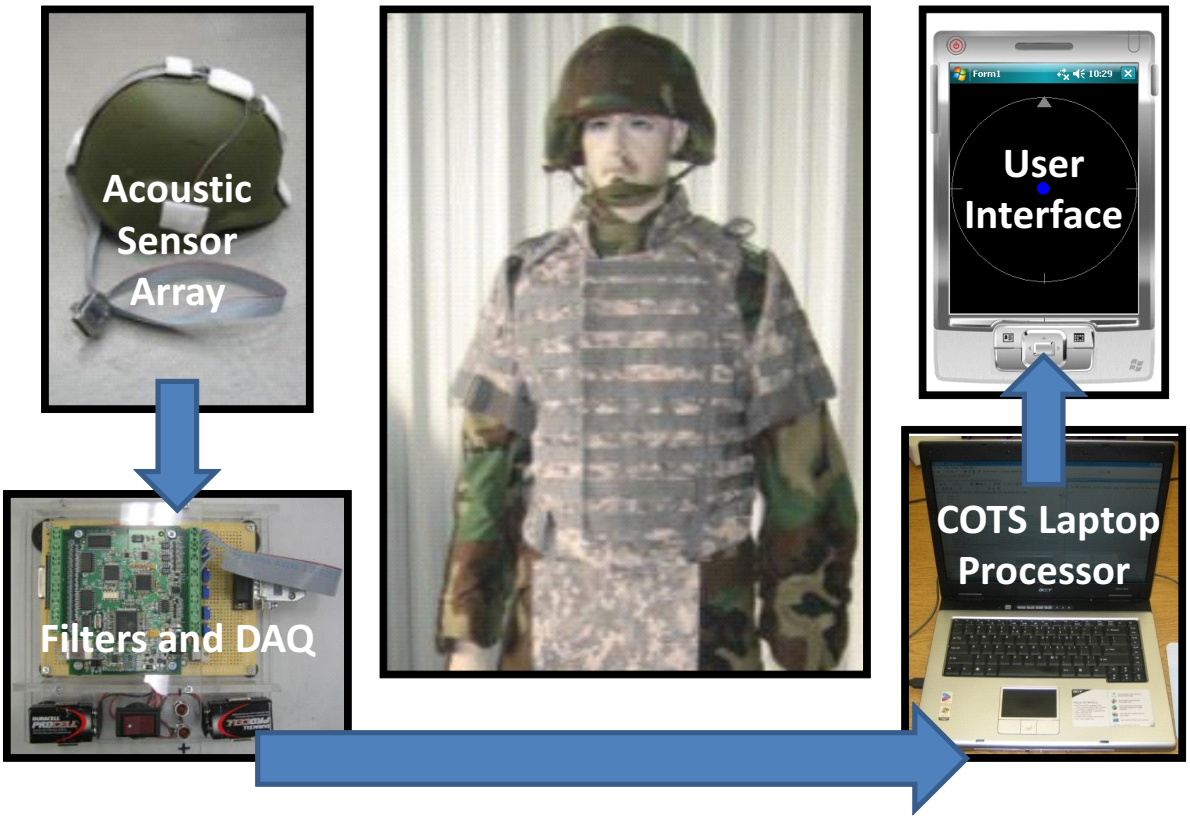


Figure 1.1 Soldier Battlespace Auditory Analyzer System (SBAAS) prototype developed by Brush Mountain Technologies under Army Phase I contract A08-T019.

by soldiers while adding negligible weight and complexity to the systems currently carried by soldiers. An example of this is that based on experience, soldiers can generally correctly identify the type of gun fired when the shooter range is greater than 100 yards. However, at ranges less than 100 yards, this ability to discern is no longer reliable; a perfect niche for the augmented performance of an acoustic array processing system. The limitations of a soldier mounted system include weight, mounting locations, power consumed, additional training required, physical durability, robustness of system algorithms, etc. all weighed against the benefit afforded the system wearer. In the case of the helmet mounted system the concern of adversely affecting the auditory cues of the wearer should be of little importance provided the system performance is trusted enough to completely supplant human perception. An additional goal was to account for the diffractive effects experienced by the array, providing an improvement in localization accuracy.

1.1.2 Headset Mounted

Placing a set of headphones on a human provides several desirable features such as hearing protection, control over the acoustic environment (via headphone speakers), and a means of communication. The shortcoming of headphones is the complete occlusion of the pinnae (the ears), disrupting auditory cues utilized by humans for sound localization [6].

As an analogy, imagine a human who has been deaf for his/her entire life and can suddenly hear. Such an event would be nothing short of a miracle, but the immediate effects would likely be confusion at the new stimulus presented to the brain and the advent of a learning process to cope with the perception of auditory stimuli. The case of headphones is not far removed from this medical miracle, as the headphone wearer is forced to completely relearn how to perceive and localize sounds. In addition, the auditory cues are not only different from our natural cues but most likely lacking in the complexity required as headphones lack complex geometric features. One caveat is that the headphone wearer is at a distinct disadvantage, because the short term nature of headphones allows only enough time for frustration before being removed. The more common scenario is for a human to hear a sound, remove the headphones to hear the

sound again for accurate localization, and then replace the headphones once adequate localization has been performed [7].

1.1.2.1 Example Application

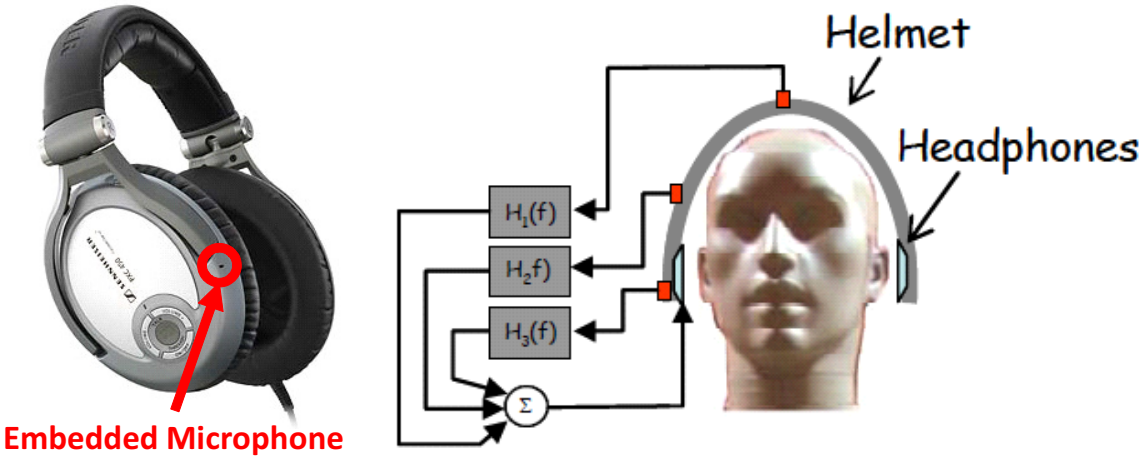
In addition to the aforementioned goals and requirements of a helmet mounted system, a headset mounted array may perform the additional function of processing and playing environmental sounds back to the wearer while preserving auditory directional cues. Commercially available headsets exist that offer sound feed through capability (e.g. Sennheiser TalkThrough technology, shown in Figure 1.2(a)), although this is meant for communication with people nearby rather than complete spectral cue restoration. Such a task is generally referred to as transparent hearing or natural hearing restoration: the goal is to recreate the auditory cues with such accuracy that the wearer feels as if he/she is not wearing headphones at all. Current transparent hearing technologies have been extensively reviewed to show that localization performance is still problematic and further study of speech intelligibility is necessary [8]. These technologies include:

- **Netherlands Organisation for Applied Scientific Research [9]** - Bronkhorst et. al. demonstrated a transparent hearing system with an equilateral triangular 3-microphone configuration on each headphone. The approach to reproducing human hearing involved "iteratively minimizing a difference measure, which was the weighted sum of log-amplitude differences and group delay differences, averaged over angle of incidence." That is to say that a direct magnitude and phase match was attempted between the array measurements and human hearing measurements. No mention of array geometry optimization is mentioned nor is any alternative array filtering strategy other than a direct match approach.

- **Adaptive Technologies Inc. [10][11]** - Goldstein, Johnson, Carneal, et. al. equipped a motorcycle helmet with a 24 element microphone array and measured the response of that array. Coupled with a phase compensation technique an 8 element subset of the original array was chosen for further testing. In the absence

of a computational model, an experimental optimization approach was used that involved testing a large set of microphones and downselecting to a smaller set. Subjective testing carried out indicated positive results in a particular operating frequency range. A schematic view of the ATI system is shown in Figure 1.2(b).

- **AuSIM Inc.** [12] - A wide range of array geometries and filtering strategies were tested by East and West coast laboratories as a transparent hearing study for the US Air Force Research Lab and Army Natick Research Lab. Biologically inspired designs such as a set of ear shaped microphone mounts were also tested for viability. The selected filtering strategy used was a least-squares magnitude (LSM) match, although this strategy was not employed in determining optimal microphone positions. Computational modeling is mentioned but no complete model, optimization method, or results are ever presented. As in the case of Adaptive Technologies Inc., the microphone position determination appears to have been performed experimentally.



(a) Sennheiser PXC 450 (b) ATI transparent hearing system

Figure 1.2 Sennheiser PXC 450 NoiseGard active noise-canceling headphones with Talk-Through and a schematic of the Adaptive Technologies Inc. transparent hearing system.

Despite the current limitations experienced by both helmet and headset mounted array systems, microphone position optimization will inevitably lead to greater flexibility

and performance in newly designed systems. The aim of this thesis is to provide a thorough analysis of performance metrics and ultimately show optimization methods that can be implemented in the design of head mounted microphone arrays.

1.2 System Overview

The following formulation is a generalized framework applicable to all arrays and is introduced here as it will have an impact on the structure of this thesis. Consider a microphone array consisting of N sensors measuring sound in an acoustic field, with an example shown in Figure 1.3. The signals from the array, $\mathbf{x}(\omega)$, are processed through a set of filters, $\mathbf{w}$, to generate an array output, $y(\omega)$,

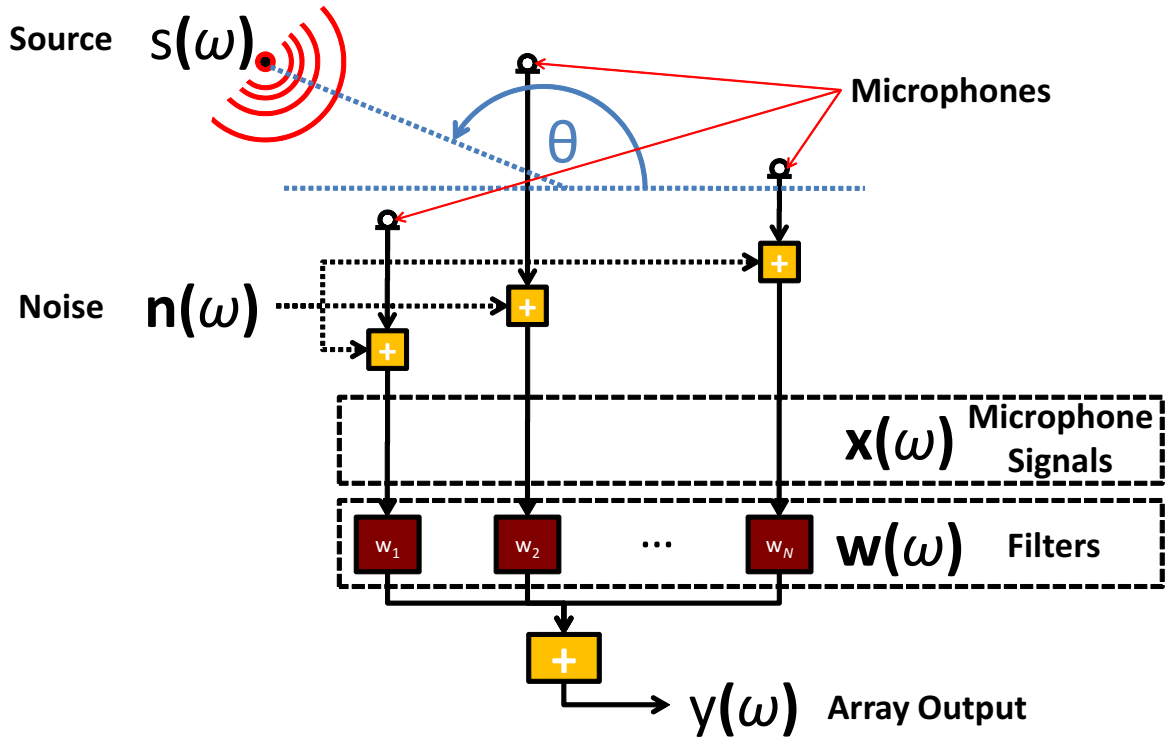


Figure 1.3 Filtering for an array with N microphones.

$$y(\omega) = \mathbf{w}^H \mathbf{x}(\omega) \tag{1.1}$$

where ω is the operating frequency and ^{H} denotes the complex conjugate transpose, or the Hermitian. Therefore, both $\mathbf{x}$ and $\mathbf{w}$ are complex column vectors of length N and y is a single complex value. In general, the microphone signals contain contributions from both "sources" and "noise." For the case given in Figure 1.3, the microphones signals, $\mathbf{x}$, are due to a single source $s(\omega)$ present at angle θ (here θ is shown as a single angle, although for the general case θ is a parameter that represents both azimuth and elevation angles) and noise, $\mathbf{n}(\omega)$,

$$\mathbf{x}(\omega) = \mathbf{d}_\theta s(\omega) + \mathbf{n}(\omega) \tag{1.2}$$

where $\mathbf{d}_\theta$ is the array manifold vector: the set of magnitude and phase differences that exist between the microphone signals for a source at angle θ.

The array manifold vector is affected by the array geometry, microphone sensitivities, source angle, frequency, and presence of any diffracting objects near the array. Each angle has a corresponding array manifold vector and the response from a number of angles can be collected into an array manifold matrix (this will be dealt with in more detail in section 2.1).

The measured noise, $\mathbf{n}(\omega)$, can take many forms [13]. In this dissertation, noise will take the form of uncorrelated noise at the sensors (i.e. noise at any one sensor is not correlated with noise at all other sensors), a single interferer (an undesirable signal originating from a single spatial location), or diffuse noise (a large collection of spatially distributed interferer's, i.e. noise at any one sensor **is** correlated with noise at all other sensors) [14]. Since diffuse noise is comprised of contributions from sources in the environment, it is dependent upon the array manifold vectors.

The optimal set of filters, $\mathbf{w}$, is dependent upon 3 things: the objective of the array application, the array manifold matrix, and noise. The set of applied filters $\mathbf{w}$ can be fixed *a priori* to achieve a specific task or can be adapted *in situ* to accommodate changes in the signal and noise.

1.3 Objectives

In order to apply the signal models presented in Equations 1.1 and 1.2 to helmet and headset mounted microphone arrays, it is necessary to calculate the manifold vectors for these complex diffracting geometries, develop filtering strategies, develop performance metrics, and optimize array geometries.

Chapter 2 presents a method for determining array manifold vectors for microphone arrays placed in the free field and on diffracting objects (i.e. calculate $\mathbf{d}$ from Equation 1.2). Simple geometries will be used to provide a basis for how array manifold vectors are affected by geometry alone, giving way to more complex diffracting geometries that approach the complexity of the human head.

Chapter 3 validates the theoretical results of simple geometries with experimental results. The process by which array manifold vectors are measured experimentally is outlined while the explanation of reference signals is found in Appendix B.

Chapter 4 provides filtering strategies (i.e. calculate $\mathbf{w}$ from Equation 1.1) that can be implemented depending upon the stationarity of the assumed noise model. For stationary noise there are conventional or data independent methods, and for non-stationary noise there are statistically optimum or adaptive methods. An additional strategy employs both data independent and adaptive methods to suppress undesired signals.

Chapter 5 develops and presents performance metrics that will be used to analyze and compare relative benefits and/or shortcomings of array designs. These metrics will also provide a way of determining the potential performance of each array design when considering microphone array applications.

Chapter 6 takes a filtering method from Chapter 4 and applies it in a headset array system. The application involves a data independent method to isolate a voice signal and an adaptive method to suppress interfering noise signals.

Chapter 7 reviews the methods by which humans localize sound as a basis for an optimized filtering strategy. Optimal microphone positions are determined for a headset mounted array for the application of transparent hearing. A near optimal array geometry

is implemented in realtime experiments to perform subjective tests and the results are presented and discussed.

The first goal of this thesis is to show the underlying theory in designing microphone arrays placed on diffracting bodies, specifically the human head. The second goal is to provide the methods that will be employed in processing the array signals for particular head mounted array applications. The third goal is to develop a framework for determining the performance potential of microphone array designs regardless of geometric complexity. The final goal is to present a method for optimizing microphone locations for the purpose of the recreation of HRTFs, allowing transparent hearing (or natural hearing restoration) to be performed.

1.4 Contributions

- **Performance of dissimilar arrays** - The magnitudes of dipole and quadrupole singular values are presented as performance metrics for comparing different array geometries.

- **Helmet mounted array** - Capturing diffractive effects within the array manifold matrix for a helmet mounted microphone array is shown with a data independent beamformer (calculated as a superdirective beamformer using a cylindrically isotropic noise field) to track tonal noise sources (in addition to the counter-sniper application currently offered in the literature).

- **Voice isolation using a headset mounted array** - Data independent beamformers are calculated (using superdirective methods) assuming a cylindrically isotropic noise field and are applied with an adaptive algorithm to remove noise interferers. The combination of filtering methods and noise assumptions shows improvement over beamforming methods.

- **Computational model of human head** - A novel computational model of a KEMAR manikin torso equipped with a Sennheiser HMEC-250 is presented and

applied in an equivalent source method simulation to determine its acoustical scattering properties. Experimental measurements are presented to prove the validity of computational results.

- **Microphone array geometry optimization method** - A phase compensation method is introduced to modify human HRTF's as the "target" response for a headset mounted microphone array,

$$D_{HRTF_{adjust},i} = |D_{HRTF,i}|e^{j\angle D_1,i} \tag{1.3}$$

A least-squares magnitude (LSM) match is proposed for recreation of interaural level difference (ILD) cues at all frequencies. The phase compensation and LSM match methods are applied to determine the optimal array geometry for a headset mounted array.

- **Demonstration of transparent hearing system** - The optimal headset array geometry is constructed and tested against the performance of occluded localization. Results verify the optimal geometry and filtering method produce statistically significant localization improvement (using a 95% confidence interval) over the occluded case.

CHAPTER 2

Array Modeling

Webster's online dictionary defines an array as an "order or arrangement, as of troops drawn up for battle," a "regular order or arrangement; series," or most aptly "an arrangement of interrelated objects or items of equipment for accomplishing a particular task [15]." Regardless of the sensor modality (in-air acoustics, radio waves, visible light, etc.), the concepts applied to an array of similar sensors are fairly universal.

This chapter will present the quantity used to describe the response of any array: the array manifold matrix. A method for calculating the array manifold matrix for free field arrays will be given and then illustrated using line and circular arrays. With the help of analytical methods, the effect of diffraction on the array manifold matrix with be shown using arrays mounted on simple diffracting shapes. The calculation of the array manifold matrix for more complex geometries will be shown using numerical methods and models ranging from a cylinder to a 3-D digital scan of a KEMAR manikin head.

An ideal point acoustic (or electromagnetic for that matter) source radiates equally in all directions (i.e. spherical spreading). At distances very close to the point source the wave-front is still curved, but for larger distances the wave-front can be approximated as a plane wave. The close distance is generally referred to as the near-field (or Fresnel region) while the far distance is termed the far-field (or Fraunhofer region). While there are many applications that exploit the properties of near-field acoustics [16][17], the far-field assumption will be made here to simplify the presentation of concepts. A commonly accepted definition for the distance r between the source and array to be considered the

far-field is

$$r = \frac{2L^2}{\lambda} \tag{2.1}$$

where L is the largest dimension of the array and λ is the wavelength of interest. Several other definitions of the far-field exist, but for the purpose of illustration the aforementioned expression will suffice. In general, array processing techniques are developed with the assumption of far-field sources, something that may cause problems if implemented in a physical design when sources occur in the near-field [18].

The array manifold matrix for an array is a collection of sensor responses to many angles of interest. The angles used are based on the particular application of the array. Typical geometries include cylindrical (i.e. azimuth plane) and spherical (i.e. all directions in 3-D space). For arrays in the free field, the manifold matrix is determined by looking at the distance between sensors and sources, or delays in the propagation of the wave across the sensors. For arrays mounted on rigid objects, such as cylinders and spheres, the wave is diffracted and more complicated expressions for the diffraction of waves across the sensors are required. As diffracting shapes become more complex, numerical models are required to determine scattered sound fields. Once a physical realization of the array has been created, experimental methods can be used to account for diffraction, sensitivities, and other dynamics of the system.

2.1 Array Manifold Matrix

The array manifold is the most complete way of objectively describing the response of an array to sources at a given frequency. In short, the array manifold, a hypersurface embedded within multidimensional space [19], is a "fingerprint" for each array. Mathematical analyses of array manifolds have been performed through the application of differential geometry [20][21][22] producing one of the most notable algorithms in array signal processing, the multiple signal classification (MUSIC) algorithm [23].

The manifold matrix is a discretization of the array manifold. Given any arbitrary array geometry having N sensors, the array manifold matrix at a frequency ω for M

far-field sources (at M angles relative to the array) can be written as

$$\mathbf{D}(\omega) = [\mathbf{d}_{\theta_1}(\omega) \quad \mathbf{d}_{\theta_2}(\omega) \quad \dots \quad \mathbf{d}_{\theta_m}(\omega) \quad \dots \quad \mathbf{d}_{\theta_M}(\omega)] \tag{2.2}$$

where each array manifold vector $\mathbf{d}_{\theta_m}(\omega)$ can be expressed as

$$\mathbf{d}_{\theta_m}(\omega) = [a_{1,m}e^{i\gamma_{1,m}} \quad a_{2,m}e^{i\gamma_{2,m}} \quad \dots \quad a_{n,m}e^{i\gamma_{n,m}} \quad \dots \quad a_{N,m}e^{i\gamma_{N,m}}]^T \tag{2.3}$$

where $\gamma_{n,m}$ and $a_{n,m}$ are the phase angle and response magnitude between the n^{th} sensor and the m^{th} source, respectively.

Combining Equations 2.2 and 2.3 yields the array manifold matrix for N sensors at M angles at a single frequency ω.

$$\mathbf{D}(\omega) = \begin{bmatrix} a_{1,1}e^{i\gamma_{1,1}} & a_{1,2}e^{i\gamma_{1,2}} & \cdots & a_{1,m}e^{i\gamma_{1,m}} & \cdots & a_{1,M}e^{i\gamma_{1,M}} \\ a_{2,1}e^{i\gamma_{2,1}} & a_{2,2}e^{i\gamma_{2,2}} & \cdots & a_{2,m}e^{i\gamma_{2,m}} & \cdots & a_{2,M}e^{i\gamma_{2,M}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n,1}e^{i\gamma_{n,1}} & a_{n,2}e^{i\gamma_{n,2}} & \cdots & a_{n,m}e^{i\gamma_{n,m}} & \cdots & a_{n,M}e^{i\gamma_{n,M}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{N,1}e^{i\gamma_{N,1}} & a_{N,2}e^{i\gamma_{N,2}} & \cdots & a_{N,m}e^{i\gamma_{N,m}} & \cdots & a_{N,M}e^{i\gamma_{N,M}} \end{bmatrix} \tag{2.4}$$

Since microphone arrays generally operate at more than one frequency, it is useful to combine array manifold matrices from multiple frequencies into a 3-dimensional matrix form, represented pictorially in Figure 2.1.

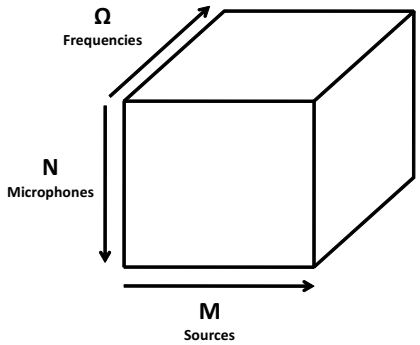


Figure 2.1 Shape of array manifold matrix for N microphones and M angles at Ω frequencies.

This illustration of the 3-dimensional array manifold matrix will be referred to in later examples.

2.2 Calculating Manifolds For Free Field Arrays

The simplest case for calculating an array manifold matrix is for microphones in the free field. For small array apertures and sources in the far-field the relative magnitude between elements can be ignored and only the phases have to be accounted for. The microphones positions (X, Y, Z) of an array of N elements can be expressed by

$$X = [x_1 \quad x_2 \quad \dots \quad x_n \quad \dots \quad x_N] \tag{2.5}$$

$$Y = [y_1 \quad y_2 \quad \dots \quad y_n \quad \dots \quad y_N] \tag{2.6}$$

$$Z = [z_1 \quad z_2 \quad \dots \quad z_n \quad \dots \quad z_N] \tag{2.7}$$

The positions (X_s, Y_s, Z_s) of M sources in the far-field can be expressed by

$$X_s = [x_1 \quad x_2 \quad \dots \quad x_m \quad \dots \quad x_M] \tag{2.8}$$

$$Y_s = [y_1 \quad y_2 \quad \dots \quad y_m \quad \dots \quad y_M] \tag{2.9}$$

$$Z_s = [z_1 \quad z_2 \quad \dots \quad z_m \quad \dots \quad z_M] \tag{2.10}$$

The distance between the n^{th} microphone and the m^{th} source is given by the Euclidean distance between points

$$d_{n,m} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2} \tag{2.11}$$

and the relationship between the phase angle $\gamma_{n,m}$ and the Euclidean distance $d_{n,m}$ is

$$\gamma_{n,m} = k d_{n,m} \tag{2.12}$$

where k is the wavenumber, defined in terms of the frequency ω and the speed of sound c,

$$k = \frac{\omega}{c} \tag{2.13}$$

At frequency ω, the array manifold matrix is

$$\mathbf{D}(\omega) = \begin{bmatrix} e^{-jkd_{1,1}} & e^{-jkd_{1,2}} & \dots & e^{-jkd_{1,m}} & \dots & e^{-jkd_{1,M}} \\ e^{-jkd_{2,1}} & e^{-jkd_{2,2}} & \dots & e^{-jkd_{2,m}} & \dots & e^{-jkd_{2,M}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ e^{-jkd_{n,1}} & e^{-jkd_{n,2}} & \dots & e^{-jkd_{n,m}} & \dots & e^{-jkd_{n,M}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ e^{-jkd_{N,1}} & e^{-jkd_{N,2}} & \dots & e^{-jkd_{N,m}} & \dots & e^{-jkd_{N,M}} \end{bmatrix} \tag{2.14}$$

Rather than represented as points, sound sources are more commonly treated as coming from a particular angle with respect to the array. This is accomplished by choosing a particular microphone (or point) as a reference for the array. For example, choosing microphone 1 as the reference point changes the term $d_{n,m}$ to $\Delta d_{n,m}$,

$$\Delta d_{n,m} = d_{n,m} - d_{1,m} \tag{2.15}$$

Without loss of generality, this changes the form of the array manifold matrix to,

$$\mathbf{D}(\omega) = \begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 \\ e^{-jk\Delta d_{2,1}} & e^{-jk\Delta d_{2,2}} & \dots & e^{-jk\Delta d_{2,m}} & \dots & e^{-jk\Delta d_{2,M}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ e^{-jk\Delta d_{n,1}} & e^{-jk\Delta d_{n,2}} & \dots & e^{-jk\Delta d_{n,m}} & \dots & e^{-jk\Delta d_{n,M}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ e^{-jk\Delta d_{N,1}} & e^{-jk\Delta d_{N,2}} & \dots & e^{-jk\Delta d_{N,m}} & \dots & e^{-jk\Delta d_{N,M}} \end{bmatrix} \tag{2.16}$$

2.2.1 Example: Line Array

A line array with equally spaced elements will be used for an illustration of the array manifold vectors and their impact on array performance. Assume a 20 microphone equally spaced line array with spacing d between elements, aligned with the x-axis as shown in Figure 2.2. For the examination of microphone phases, configurations between endfire ($\theta = 0°$) and broadside ($\theta = 90°$) will be used, with Figure 2.2 defining the azimuth and elevation angles between the source and the array. The following figures represent the

plots of the columns of the 3-dimensional array manifold matrix at 3 different frequencies, illustrated in Figure 2.3.

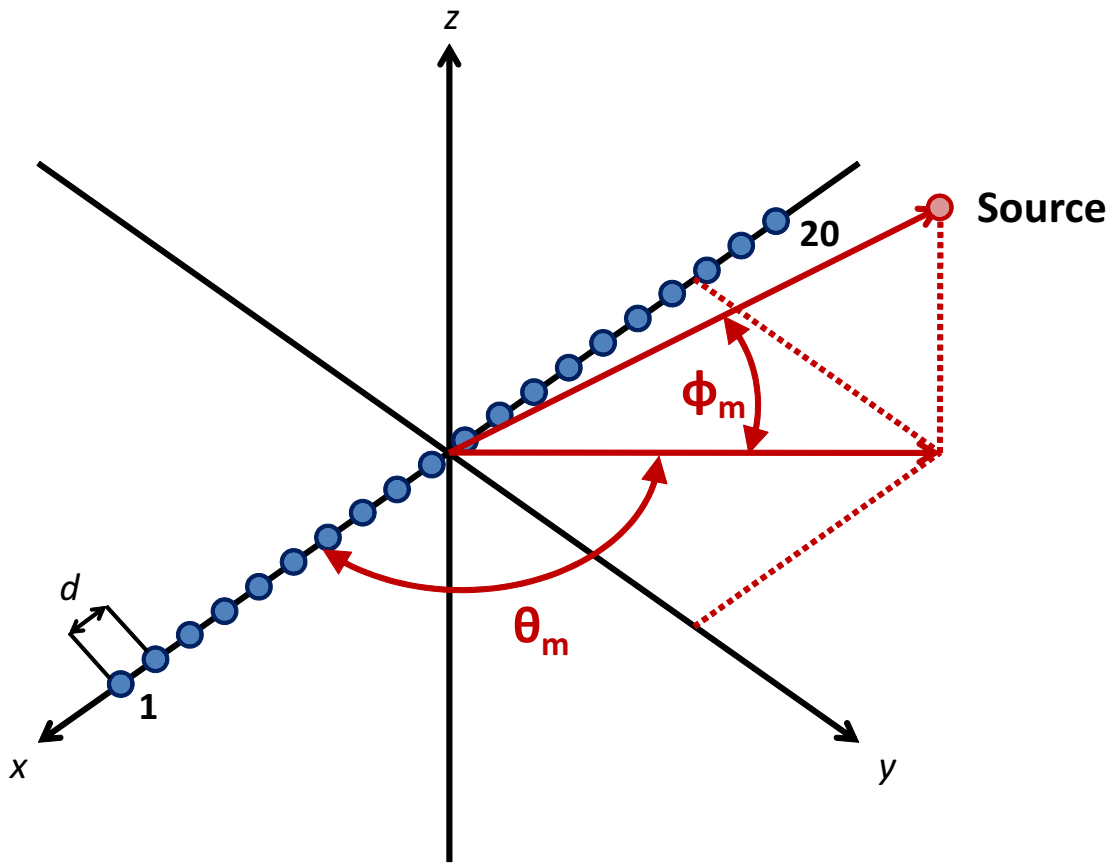


Figure 2.2 Geometry for a line array with 20 microphones and a source at azimuth angle θ_m and elevation angle ϕ_m.

Figure 2.4 shows the microphones phases for $kd = \pi/1000$, $kd = \pi/10$, and $kd = 10\pi$ sources at azimuth angles between endfire and broadside configurations. All phases are taken relative to the 11th microphone, which is near the center of the array, hence the presence of 0° phase for the 11th microphone in every case. In Figure 2.4(a) the wavelength is very long compared to the spacing between microphones, causing the array to measure very similar phases at all microphones regardless of source angle. To tie this example to linear algebra, Figure 2.4(a) corresponds to a set of nearly linearly dependent array manifold vectors, preventing discrimination of source angle because all source angles have very similar array responses. In Figure 2.4(c) the wavelength is very short compared

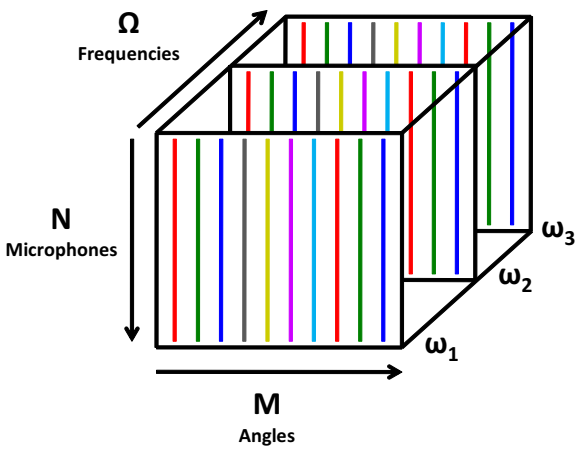
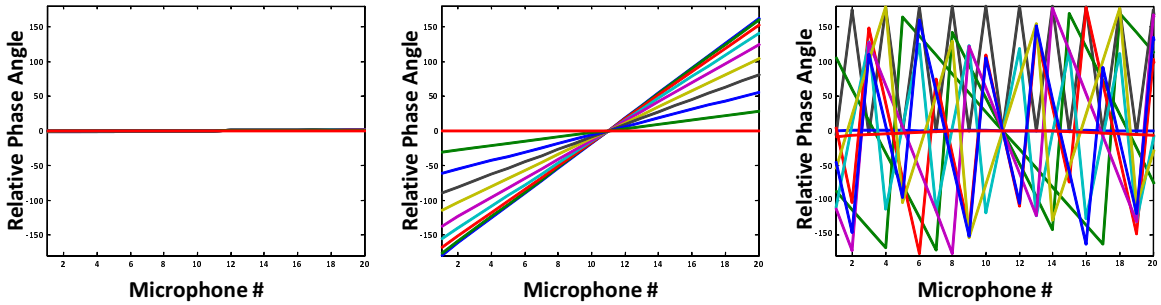


Figure 2.3 Illustration of array manifold matrix columns plotted in Figures 2.4, 2.5, 2.7, and 2.8.

to the spacing between microphones, meaning that multiple wavelengths may fit within the microphone spacing. The problem in Figure 2.4(c) is that the phase is indeterminate because of the $(0, 2\pi]$ limits on phase measurement.

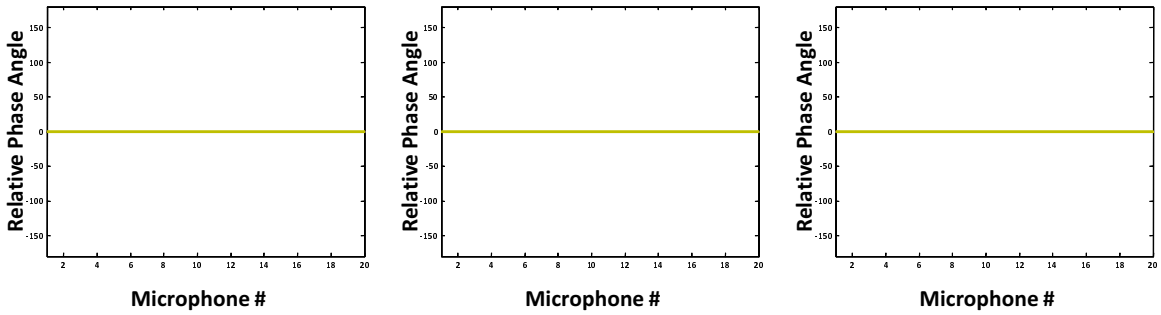


(a) $kd = \pi/1000$ (b) $kd = \pi/10$ (c) $kd = 10\pi$

Figure 2.4 Relative phase between elements of a 20 microphone line array for different kd values and source positions of $\theta = 0°$, $10°$, $20°$, $30°$, $40°$, $50°$, $60°$, $70°$, $80°$, and $90°$ at $\phi = 0°$.

A visual representation of the array manifold matrix can also show why a line array suffers from a "cone of confusion," or a cone shaped spatial region where all sources within that region produce the same array response (assuming omnidirectional sensors). Figure 2.5 shows the relative phases for $kd = \pi/1000$, $kd = \pi/10$, and $kd = 10\pi$ sources for a broadside configuration and all elevation angles. The flat line feature of Figure 2.5

is indicative of a line array in a broadside configuration, so that should not be alarming. The interesting feature to note is that sources at ALL elevation angles have IDENTICAL relative phases at the microphones. That is to say that the array manifold vectors represented by these phases are just repetitions.



(a) $kd = \pi/1000$ (b) $kd = \pi/10$ (c) $kd = 10\pi$

Figure 2.5 Relative phase between elements of a 20 microphone line array for different kd values and source positions of $\theta = 90°$ at $\phi = 0°$, $10°$, $20°$, $30°$, $40°$, $50°$, $60°$, $70°$, $80°$, and $90°$.

2.2.2 Example: Circular Array

Using the same sets of angles used for Figure 2.4 and Figure 2.5 to examine the broadside response of a 20 element circular array with equal element spacing d, shown in Figure 2.6, yields the results shown in Figure 2.7 and Figure 2.8, respectively. Figure 2.7 shows much the same results exhibited from the line array case in Figure 2.4, albeit with a different shape to the phase angles between elements: curved rather than linear. The same wavelength issues appear where excessively long (Figure 2.7(a)) or excessively short (Figure 2.7(c)) wavelengths with respect to the array aperture cause either identical or indeterminable phase differences, respectively.

The performance gained by rearranging the sensors into a circular geometry is discrimination between sources at different elevations. This occurs as the array becomes 2-dimensional. Figure 2.8 shows the same analysis as Figure 2.5, but this time the relative phase differences are not flat and identical, but rather similar to those seen in Figure 2.7, where azimuth discrimination was demonstrated. It should be noted that top-down

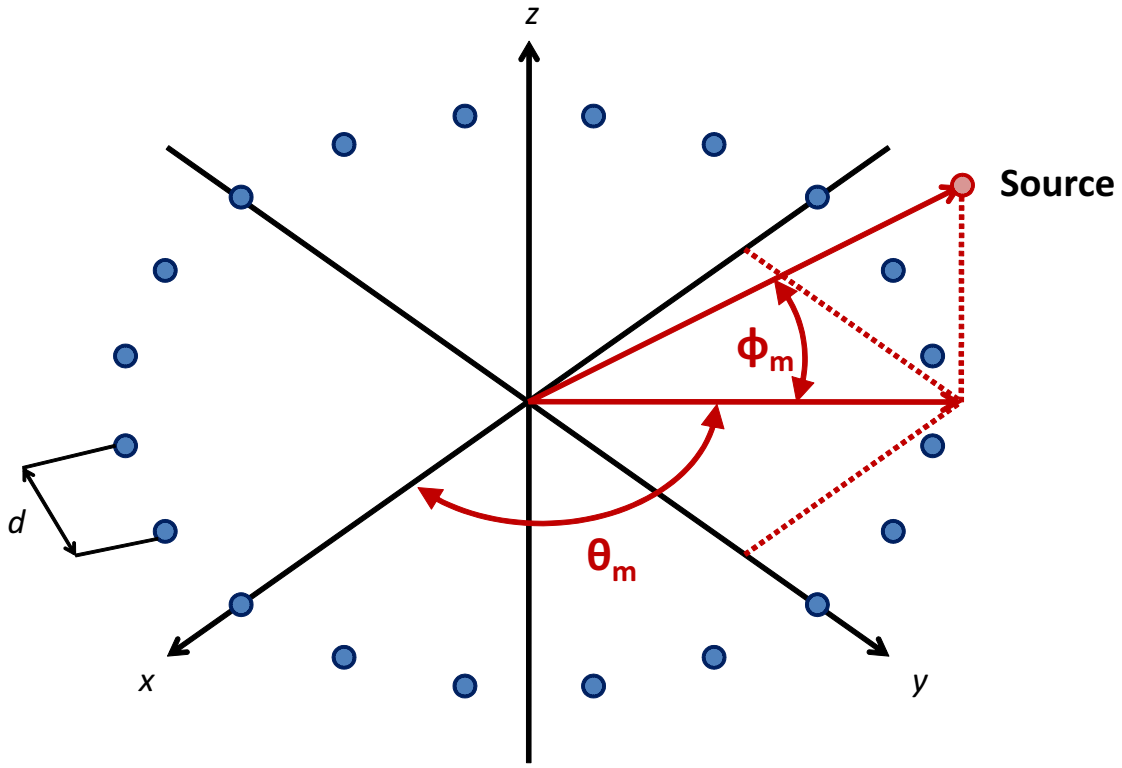
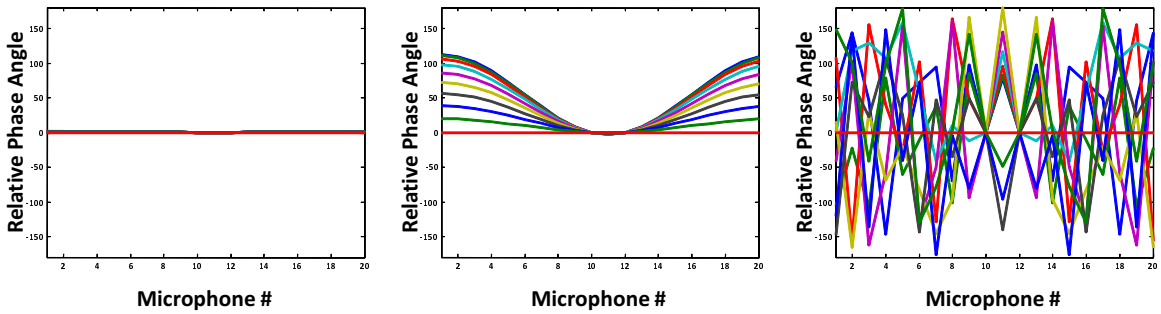


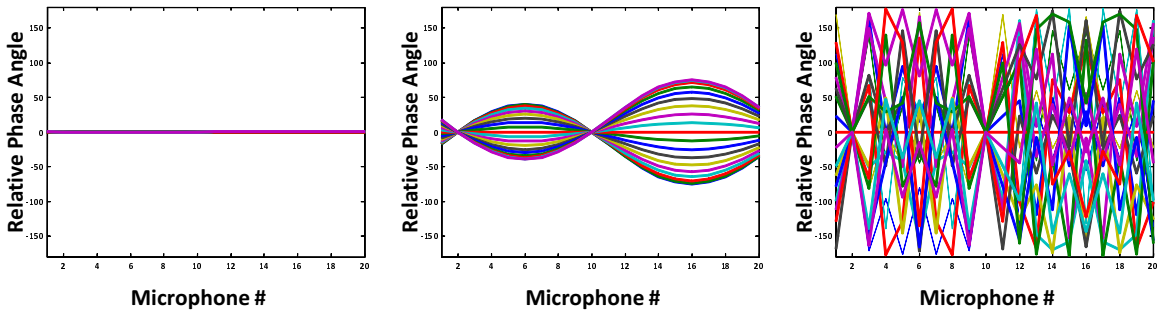
Figure 2.6 Geometry for a circular array with 20 microphones and a source at azimuth angle θ_m and elevation angle ϕ_m.



(a) $kd = \pi/1000$ (b) $kd = \pi/10$ (c) $kd = 10\pi$

Figure 2.7 Relative phase between elements of a 20 microphone circular array for different kd values and source positions of $\theta = 0°$, $10°$, $20°$, $30°$, $40°$, $50°$, $60°$, $70°$, $80°$, and $90°$ at $\phi = 0°$.

ambiguity still exists for this planar array design (i.e. sources at $+\phi$ produces the same response as sources at $-\phi$ degrees).



(a) $kd = \pi/1000$ (b) $kd = \pi/10$ (c) $kd = 10\pi$

Figure 2.8 Relative phase between elements of a 20 microphone circular array for different kd values and source positions of $\theta = 90°$ at $\phi = 0°$, $10°$, $20°$, $30°$, $40°$, $50°$, $60°$, $70°$, $80°$, and $90°$.

2.3 Calculating Manifolds For Arrays Mounted On Simple Diffracting Shapes

Finding the array manifold for a microphone array mounted on a diffracting object is more complicated than for a free field array. While the terms in the array manifold vectors for a free field array consist of only phase differences due to propagation distances (see Equation 2.16), the vectors for diffracting arrays contain phase **and** magnitude differences based on the acoustical scattering properties of the mounting body. To determine the diffracted sound field from a rigid object the incident and scattered waves add together to a zero normal particle velocity at the surface of the rigid object [24].

The form of the plane wave pressure field, neglecting time dependence, is the same regardless of the rigid object shape,

$$p_p = P_0[e^{ik_x x} e^{ik_y y} e^{ik_z z}] \tag{2.17}$$

where P_0 is the pressure amplitude and k_x,k_y, and k_z are the directional wavenumbers, defined by

$$k_x = k\cos\theta\cos\phi \tag{2.18}$$

$$k_y = k\sin\theta\cos\phi \tag{2.19}$$

$$k_z = k\sin\phi \tag{2.20}$$

For a wave assumed to be traveling along the x-axis the formulation is simplified to

$$p_p = P_0 e^{ikx} \tag{2.21}$$

The form of the scattered pressure field will depend upon the shape of the rigid object, and the addition of the two pressure fields will be the total diffracted sound field. The diffraction by rigid cylinders and spheres can be calculated analytically and has been verified by experiments [25].

2.3.1 Cylinder

The manifold for a circular array mounted on an infinite cylinder of radius a is determined by examining the pressure at the surface of the cylinder. The total pressure field on the surface of a cylinder due to a plane wave with unit pressure magnitude in the azimuth $(x-y)$ plane has the form [26]

$$p(r,\psi_{n,m}) = \sum_{q=0}^{\infty} \delta_q (-i)^q \cos\psi_{n,m}[J_q(ka) + D_q H_q(ka)] \tag{2.22}$$

where,

$$\delta_q = \begin{cases} 1 & q = 0 \\ 2 & q > 0 \end{cases} \tag{2.23}$$

$$D_q = -\frac{(-iG + \frac{q}{ka})J_q(ka) - J_{q+1}(ka)}{(-iG + \frac{q}{ka})H_q(ka) - H_{q+1}(ka)} \tag{2.24}$$

where G is the surface admittance of the cylinder (assumed to be zero for a rigid condition), k is the wavenumber of the incoming sound wave, J_q and H_q are Bessel and

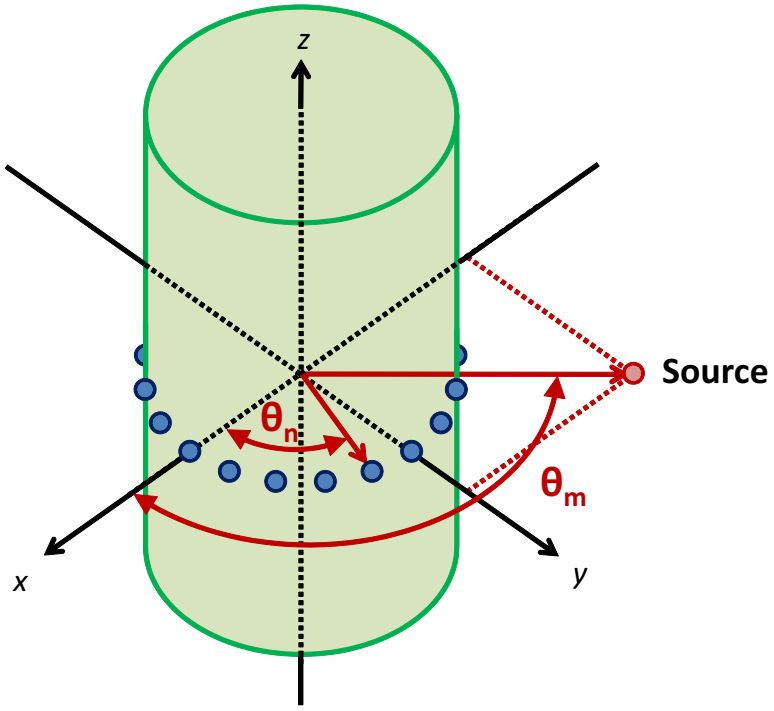


Figure 2.9 Angle between source direction and microphone direction.

Hankel functions, respectively, and $\psi_{n,m}$ is the azimuth angle between the m^{th} source direction and the n^{th} microphone location, defined as

$$\psi_{n,m} = \theta_n - \theta_m \tag{2.25}$$

illustrated in Figure 2.9. A complete derivation of the scattered sound field is found in the literature [27][24].

The presence of a rigid cylinder as the mounting body for the array produces an array manifold vector composed of scattering terms for every microphone response,

$$\mathbf{d}_{\theta_m} = \begin{bmatrix} \sum_{q=0}^{\infty} \delta_q (-i)^q \cos\psi_{1,m} [J_q(ka) + D_q H_q(ka)] \\ \sum_{q=0}^{\infty} \delta_q (-i)^q \cos\psi_{2,m} [J_q(ka) + D_q H_q(ka)] \\ \vdots \\ \sum_{q=0}^{\infty} \delta_q (-i)^q \cos\psi_{n,m} [J_q(ka) + D_q H_q(ka)] \\ \vdots \\ \sum_{q=0}^{\infty} \delta_q (-i)^q \cos\psi_{N,m} [J_q(ka) + D_q H_q(ka)] \end{bmatrix} \tag{2.26}$$

where θ_m is the azimuth angle from the m^{th} source to the center of the array and all other terms have been previously defined. To calculate the response of each microphone, a few terms in a finite sum can be used to closely approximate the infinite sum.

As a brief comparison of the impact a diffracting body has on the microphone array response, Figures 2.10 and 2.11 show the phase and magnitude fields, respectively, for (a) free field and (b) diffracting circular arrays. Needless to say, the presence of a rigid body within the circular array greatly increases the relative phase differences between microphones at low frequencies, helping mitigate the lack of performance at relatively long wavelengths [28][29][30].

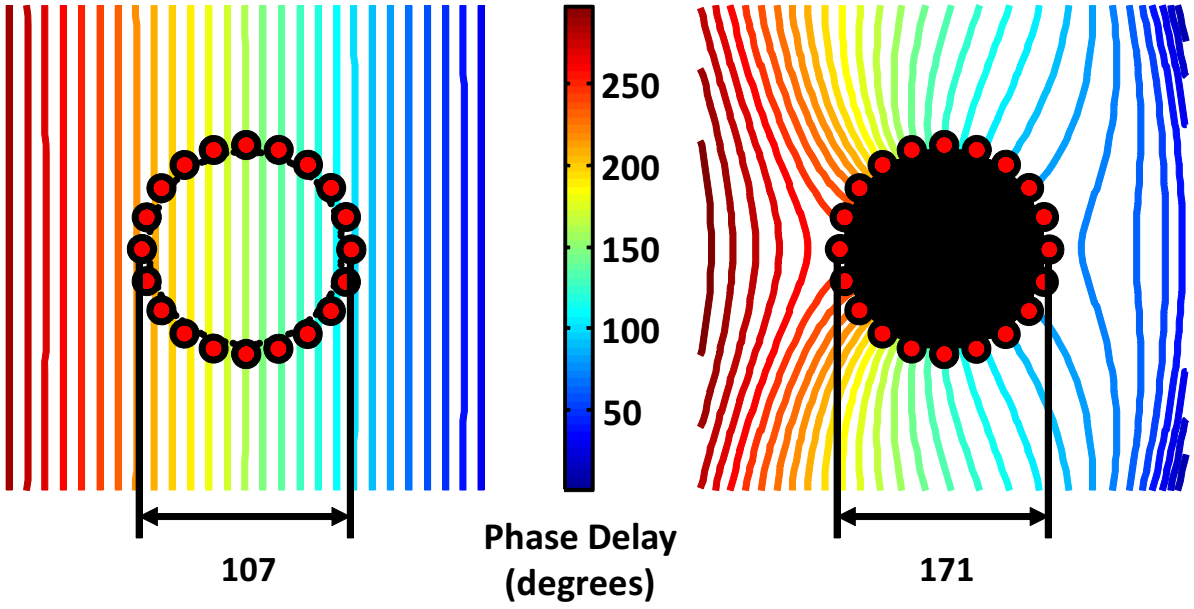


Figure 2.10 Phase fields for circular arrays (left) in the free field and (right) mounted on a rigid infinite cylinder for $ka = 0.6\pi$. The maximum phase difference between front and rear microphones is labeled in degrees for each case.

Another way to view the magnitude and phase differences caused by the presence of a rigid cylinder is in terms of the response of the n^{th} microphone to sources at all M angles. Whereas the columns of the 3-dimensional array manifold matrix correspond to the response of all microphones at each angle (shown in Figure 2.7), the rows of that matrix correspond to the response of each microphone at all angles (shown in Figure 2.12). The magnitude and phase responses of a microphone at $\theta = 0°$ to sources at all

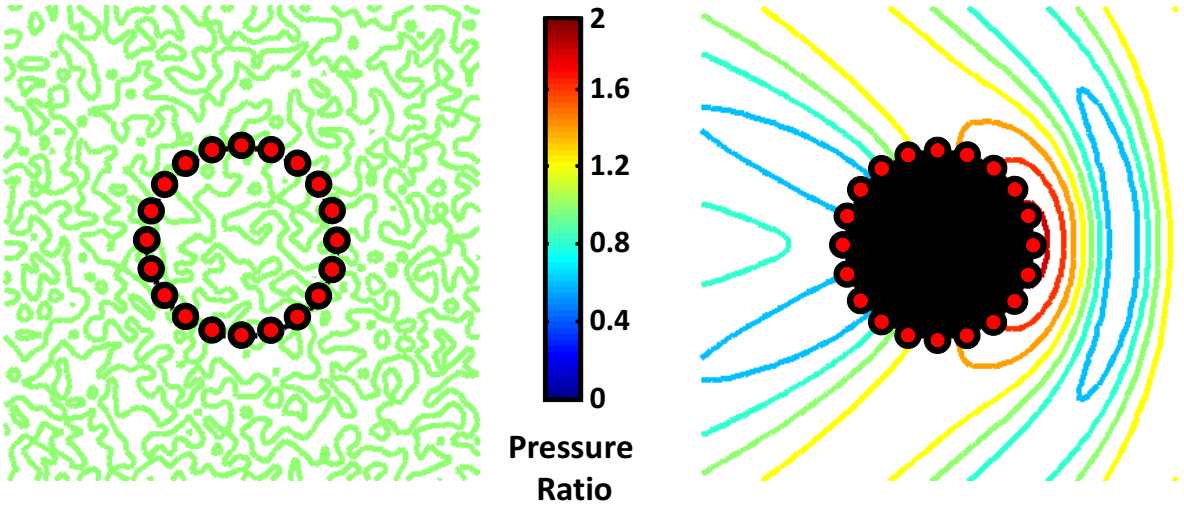


Figure 2.11 Magnitude fields for circular arrays (left) in the free field and (right) mounted on a rigid infinite cylinder for $ka = 0.6\pi$. The pressure ratio is based on the magnitude of the incoming plane wave.

azimuth angles is shown in Figure 2.13, both in the free field and on the surface of a rigid cylinder. In Figure 2.13(a), note the drastic disparity in magnitude when the microphone is facing the source compared to facing away from the source for the diffracting case. In Figure 2.13(b) the phase difference between the diffracting and free field cases occurs at all angles, but extraordinarily so when the microphone is facing away from the source.

2.3.2 Sphere

Similar to the circular array case, the manifold matrix for a spherical array mounted on a rigid sphere of radius a is determined by examining the pressure at the surface of the sphere. The total pressure field on the surface of the sphere due to a plane wave with unit pressure magnitude has the form [26]

$$p(r, \psi_{n,m}) = \sum_{q=0}^{\infty} \delta_q (-i)^q \cos\psi_{n,m} [j_q(ka) + D_q h_q(ka)] \tag{2.27}$$

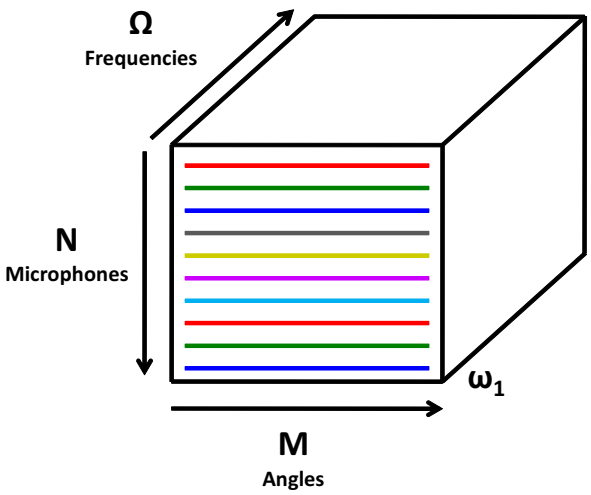
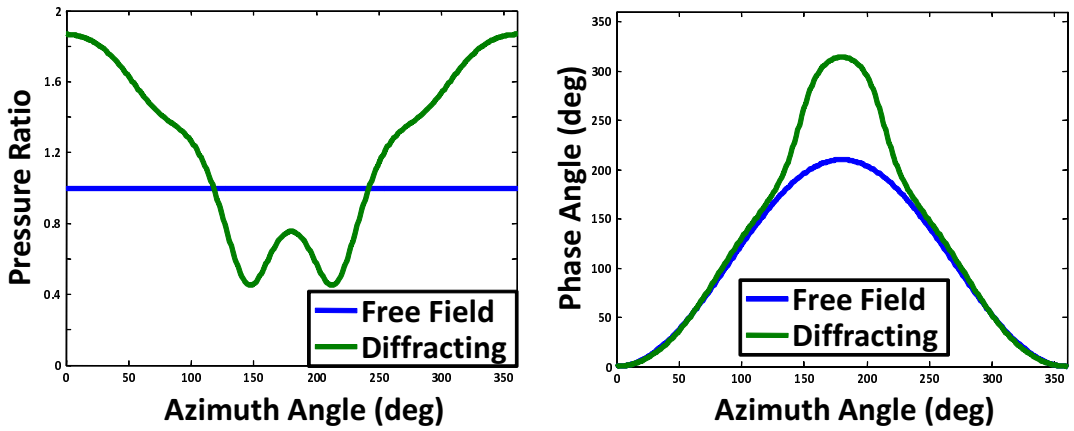


Figure 2.12 Illustration of array manifold matrix rows plotted in Figures .



(a) Magnitude (b) Phase

Figure 2.13 Response of a microphone at $\theta = 0°$ to sound sources at all azimuth angles in the free field and on a rigid cylinder for $ka = 0.6\pi$.

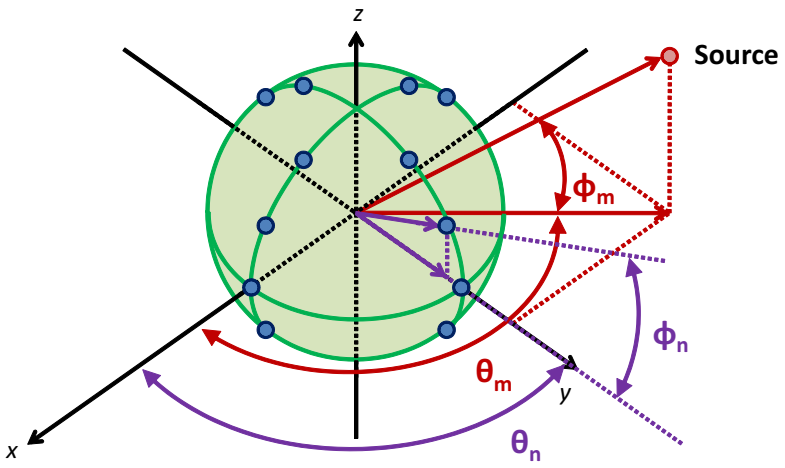


Figure 2.14 Angles for source and microphone directions for a spherical array.

where,

$$\delta_q = 2q+1 \tag{2.28}$$

$$D_q = -\frac{(-iG+\frac{q}{ka})j_q(ka) - j_{q+1}(ka)}{(-iG+\frac{q}{ka})h_q(ka) - h_{q+1}(ka)} \tag{2.29}$$

where P_q is a Legendre function of order q , G is the surface admittance of the sphere (assumed to be zero for a rigid condition), k is the wavenumber of the incoming sound wave, j_q and h_q are spherical Bessel and Hankel functions, respectively, and $\psi_{n,m}$ is the spherical angle between the source direction and the measurement location,

$$\psi_{n,m} = \arccos[(\cos\theta_n - \theta_m)\cos\phi_n\cos\phi_m + \sin\phi_n\sin\phi_m] \tag{2.30}$$

where θ_n and ϕ_n are the azimuth and elevation angles of the n^{th} microphone, respectively, and θ_m and ϕ_m are the azimuth and elevation angles of the m^{th} source, respectively, illustrated in Figure 2.14. Again, a complete derivation of the total sound field is found in the literature [27][31].

Alternatively, spherical diffraction theory is often represented in terms of spherical harmonics. Expressing the total pressure field on the surface of the sphere using spherical harmonics yields

$$p(ka,\phi,\theta) = \sum_{s=0}^{\infty}\sum_{q=-s}^{s} 4\pi i^s[j_s(ka) - \frac{j_s'(ka)}{h_s'(ka)}h_s(ka)]Y_s^{q^*}(\phi,\theta)Y_s^q(\phi,\theta) \tag{2.31}$$

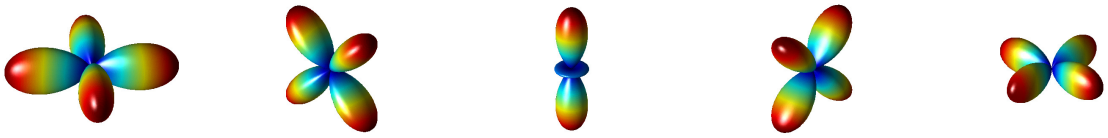
where the spherical harmonics are given by

$$Y_s^q(\phi,\theta) = \sqrt{\frac{2s+1}{4\pi}\frac{(s-q)!}{(s+q)!}} P_s^q(\cos\phi) e^{iq\theta} \tag{2.32}$$

Substituting the spherical harmonics back into Eq. 2.31 and simplifying will result in the original total pressure field given in Eq. 2.27. The spherical harmonics represent the variation in the pressure field with respect to the elevation (ϕ) and azimuth (θ) angles, while the remaining terms in the total pressure field express variation with respect to radius and wavenumber. The assumption here is that microphones are measuring pressure on the surface ($r = a$).

Spherical harmonics are solutions of Laplace's equation in spherical coordinates [31]. These harmonics represent the ways in which a sphere can radiate, with different orders having different radiation efficiencies at any given frequency. For example, the zeroth order spherical harmonic is a monopole, the first order spherical harmonics are a set of 3 dipoles, the second order spherical harmonics are a set of 5 quadrupoles, etc. At each frequency, the total pressure field at the surface of the sphere is expressed in terms of the radiation contributions from a source and the rigid sphere. The spherical harmonics are the radiation modes of a sphere, meaning that the previous equations represent a decomposition of the sound field into the various harmonics that correspond to particular frequency ranges. At low frequencies the dominant radiation mode is a monopole, or a uniform expansion and contraction of the entire sphere. As frequency is increased, the radiation efficiencies of higher order spherical harmonics (dipoles, quadrupoles, etc.) increase and contribute to the radiated sound field in greater proportions. Any sound field can be expressed as some combination of spherical harmonics. It so happens that when spherical array geometry is chosen, the scattered sound field is a pure combination of spherical harmonics. Figure 2.15 shows the spherical harmonics of the first 3 orders (s = 0, 1, 2).

(a) $s = 0$ (b) $s = 1$



(c) $s = 2$

Figure 2.15 Spherical harmonics of the first 3 orders.

Incorporating the aforementioned expressions for spherical diffraction into the array manifold vector for an array on a rigid sphere gives

$$\mathbf{d}_{\theta_m,\phi_m} = \begin{bmatrix} \sum_{q=0}^{\infty}(2q+1)(-i)^q P_q(\cos\psi_{1,m})[j_q(kr) + D_q h_q(kr)] \\ \sum_{q=0}^{\infty}(2q+1)(-i)^q P_q(\cos\psi_{2,m})[j_q(kr) + D_q h_q(kr)] \\ \vdots \\ \sum_{q=0}^{\infty}(2q+1)(-i)^q P_q(\cos\psi_{n,m})[j_q(kr) + D_q h_q(kr)] \\ \vdots \\ \sum_{q=0}^{\infty}(2q+1)(-i)^q P_q(\cos\psi_{N,m})[j_q(kr) + D_q h_q(kr)] \end{bmatrix} \tag{2.33}$$

where θ_m and ϕ_m are the azimuth and elevation angles from the source to the center of the array, $\psi_{n,m}$ is the angle between the m^{th} source and n^{th} microphone directions, and all other terms have been previously defined. Again, a finite sum comprised of a few terms can be used to approximate the infinite sums shown in Equation 2.33.

2.4 Calculating Manifolds For Arrays Mounted On Complex Diffracting Shapes

For diffracting shapes that may not have closed form analytical expressions for acoustical scattering, a computational model for diffraction is necessary. The method described

is a discrete boundary element method based on superposition of acoustic waves, referred to as the equivalent source method [32][33].

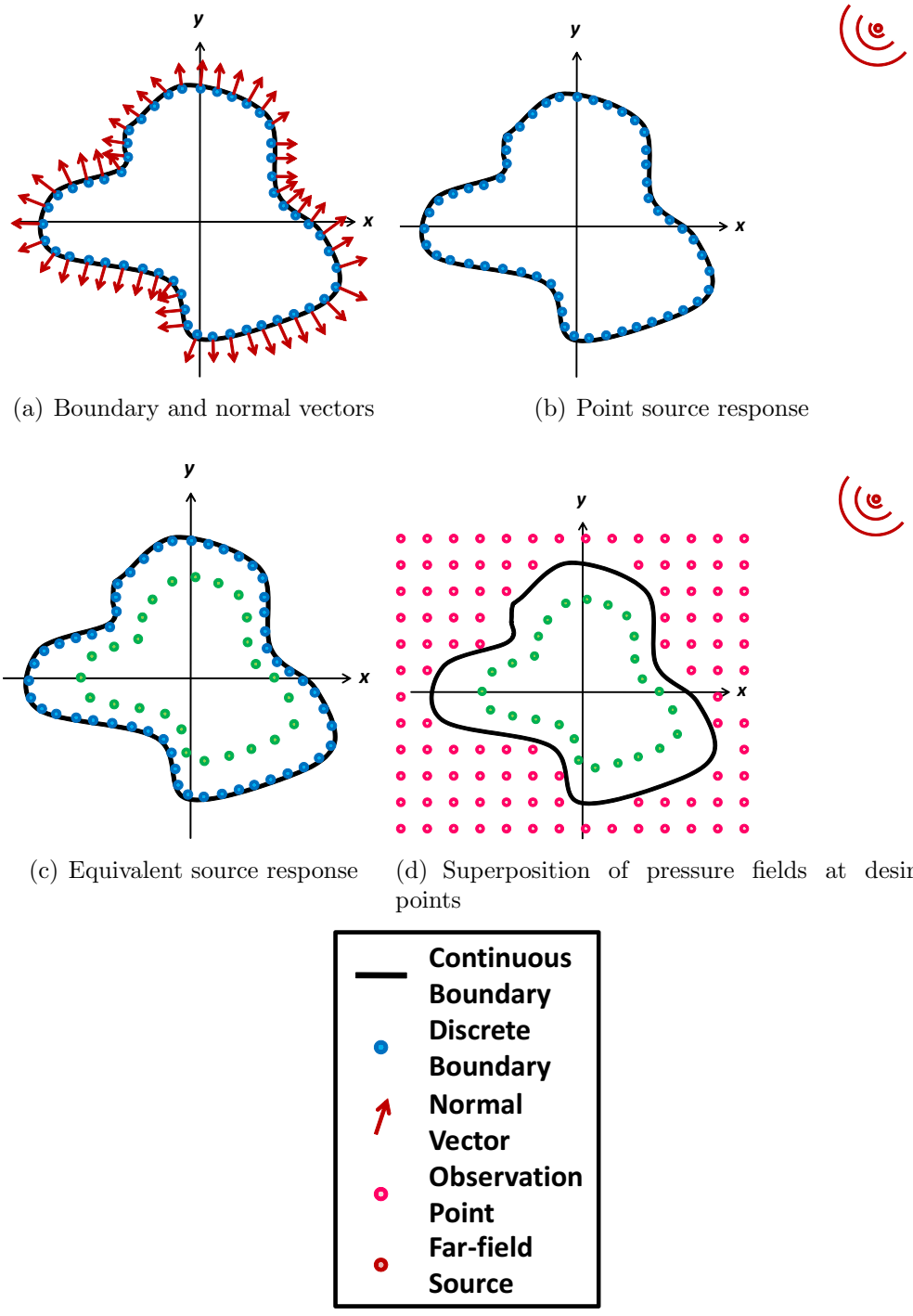
2.4.1 Cylinder

In order to model the scattering from a rigid surface, it is first necessary to describe the surface using a discrete number of points. The pressure and normal particle velocity of the surface is then calculated for a source of interest assuming a free field, shown in Figure 2.16(b) as a 2-D boundary and a point source. Next a set of "equivalent sources" is created within the interior of the rigid boundary and a least squares fit is performed to determine how these equivalent sources must be driven to achieve a rigid boundary, i.e. drive the normal velocity at the boundary to zero, shown in Figure 2.16(c). Finally, the superposition of the point source and equivalent source pressure fields is performed at points of interest, shown in Figure 2.16(d) as a grid in the $x - y$ plane.

Results from analytical expressions for the sound fields around cylinders and spheres will be used to verify the accuracy of the equivalent source method. Figure 2.17 shows the acoustical scattering of a 1000Hz plane wave caused by a 0.1016m radius, 1m tall, cylinder ($ka = 1.86$) as calculated by the equivalent source numerical method and a 0.1016m radius infinite cylinder calculated from the analytical expressions in Equations 2.22 and 2.23. At low frequencies, where the wavelength is much larger than the diameter of the sphere, the presence of the cylinder has very little impact on the sound field. At higher frequencies, where the wavelength is comparable to the cylinder diameter, the cylinder begins to reflect sound back toward the incident direction and cause a more than double of magnitude at the front of the cylinder and shadowing at the rear of the cylinder [34].

2.4.2 Sphere

A similar comparison can be made with a sphere. Figure 2.18 shows the acoustical scattering of a 1000Hz plane wave caused by a 0.1016m radius sphere ($ka = 1.86$) as



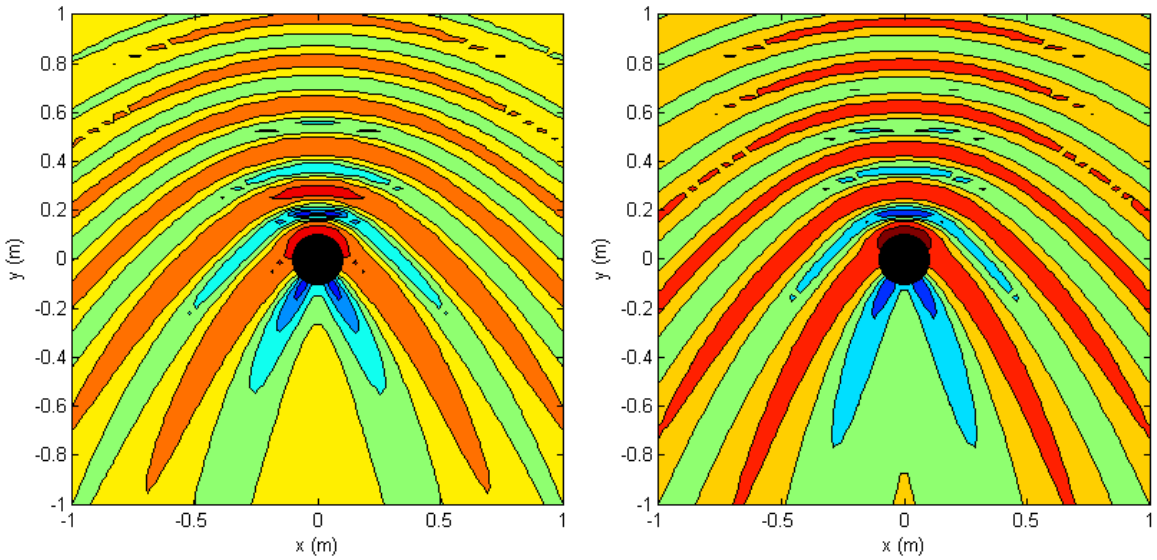
(a) Boundary and normal vectors

(b) Point source response

(c) Equivalent source response

(d) Superposition of pressure fields at desired points

Figure 2.16 Steps of equivalent source method.



(a) Numerical

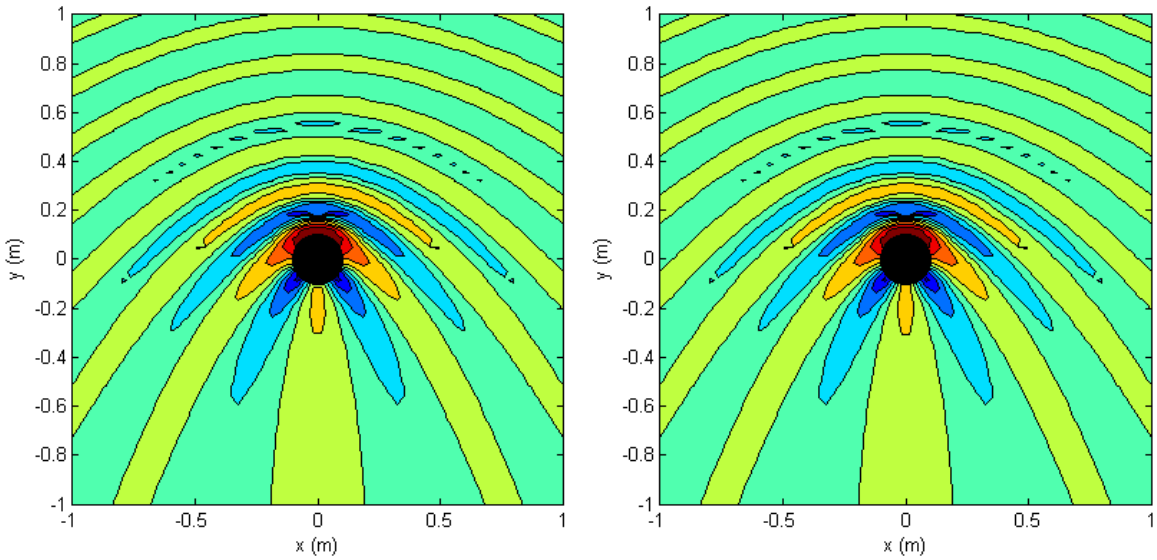
(b) Analytical

Figure 2.17 Acoustical scattering of a plane wave caused by a cylinder for $ka = 1.86$.

calculated by the equivalent source numerical method and the analytical expressions from Equations 2.27 and 2.28. The comparison is clearly next to exact.

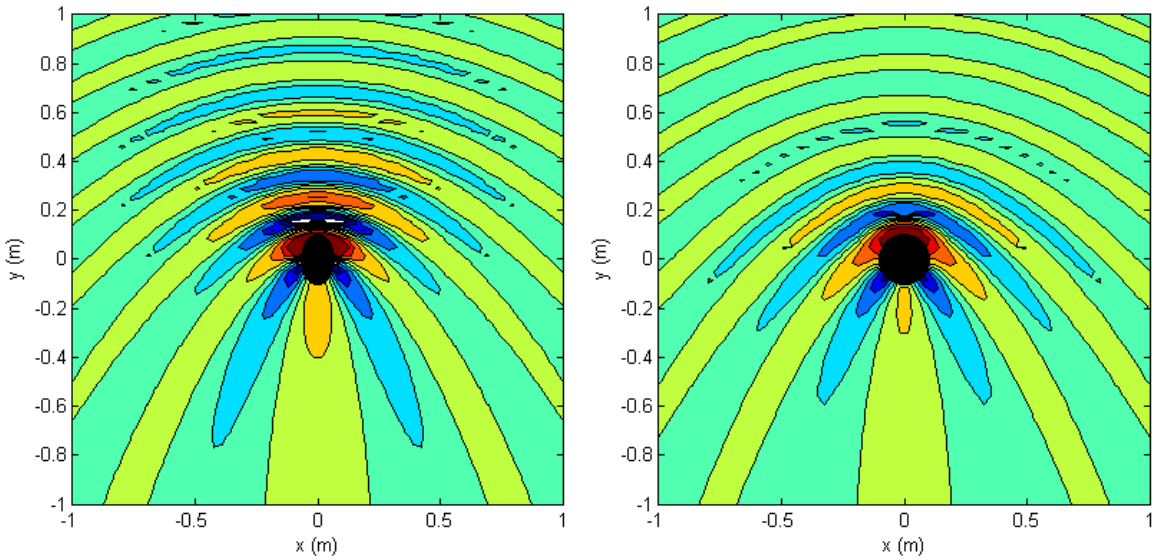
2.4.3 Prolate Ellipsoidal Approximation of Head

The human head is generally approximated by a sphere for simplicity in analyzing diffraction. However, the shape of the human head is more accurately represented by a spheroid whose axis lengths are not equal. Figure 2.19 shows the sound field caused by a 1000Hz sound wave scattering around a prolate spheroid at its middle height ($a = 0.066m$, $b = c = 0.1016m$) and the comparison to a sphere of radius 0.1016m ($ka = 1.86$). The most noticeable change in the sound field occurs at the front of the spheroid (in the direction of the sound source). The single frontal zone of increased pressure in the spherical case has split in two and moved toward the sides in the prolate spheroid case. Based on its closer approximation to the human head than a sphere, some research has explored an ellipsoidal (prolate spheroids) head model to study aspects of human hearing [35].



(a) Numerical (b) Analytical

Figure 2.18 Acoustical scattering of a 1000 Hz plane wave caused by a sphere for $ka = 1.86$.



(a) Prolate Spheroid (b) Rigid Sphere

Figure 2.19 Acoustical scattering of a 1000 Hz plane wave caused by spheroids.

2.4.4 Head, Headphones, Helmet

Several equivalent source models were developed to approximate a headphone array on the human head. Each model added more diffracting features and detail to the model to determine their effects on the response of the array. Following the simplification found in the literature, a spherical head was first used to model the headphone array. The first level of detail added was a prolate spheroid head with half prolate spheroids located in the positions of the ears as headphones. Next, a neck and torso were approximated by a cylinder and oblate spheroid, respectively. Finally, the KEMAR manikin was scanned to create a digital set of 3-dimensional points for an extremely close match to reality.

2.4.5 Spherical Head

Initially, a computer model was developed to treat the microphone array on the headset as if it were 8 microphones mounted on the side of a human head, represented as a sphere. Since several headset microphone arrays have already been characterized and measured in the past, information is available for validating the computer model that has been constructed. Figure 2.20 shows this early spherical approximation for the head, with the boundary elements shown in blue, microphone a shown in green, and microphone b shown in red. Figure 2.21 shows the first computer model results using microphone a, given as the response of a single microphone in each direction at each frequency, normalized to the maximum response of the microphone. Figure 2.22 shows the same results as Figure 2.21, but for microphone b.

2.4.6 Ellipsoidal Head

The second iteration of the computer model involved taking the head from a spherical shape to a prolate spheroid shape, where the height and depth of the head have identical axis lengths, but the width of the head (from ear to ear) has a more narrow axis length. In addition, the dimensions of the headphones were measured and their boundaries were added to the computer model on each side of the head where the ears would be located.

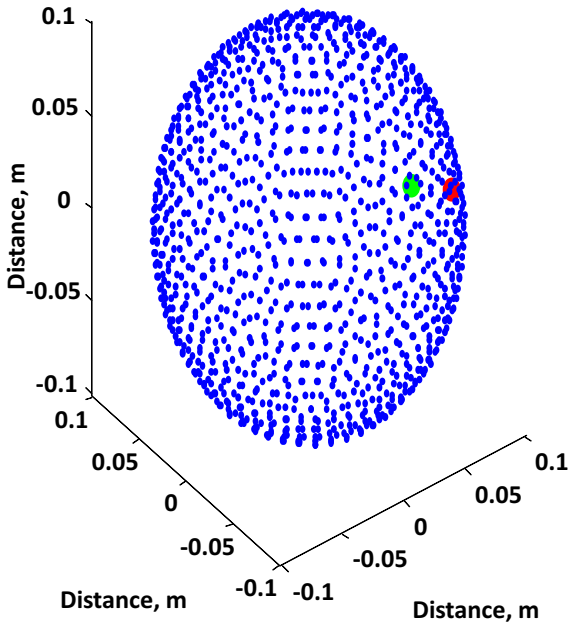


Figure 2.20 Spherical approximation for position of headphone microphones a (green) and b (red).

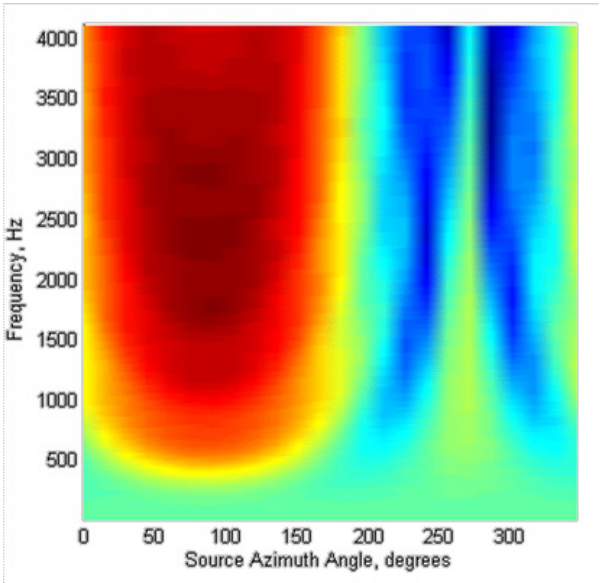


Figure 2.21 Spherical approximation results for microphone a.

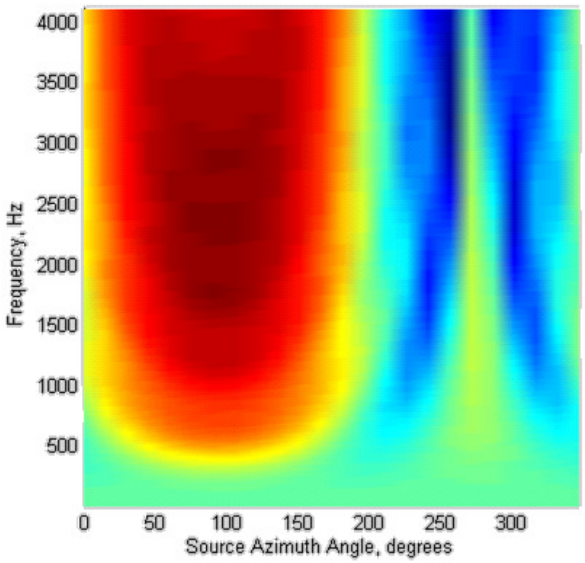


Figure 2.22 Spherical approximation results for microphone *b*.

Figure 2.23 shows this second approximation for the head and headset computer model, with the head boundary elements once again shown in blue, microphone *a* shown in green, and microphone *b* in red. Figure 2.24 shows the second computer model, given for microphone *a* the same way as Figure 2.21. Also, Figure 2.25 shows the same results as Figure 2.24, but for microphone *b*.

2.4.7 Ellipsoidal Head with Torso

Additional refinements to the second computer model were made to incorporate reflections and reverberations from the neck and shoulders, and so boundary elements for the neck and shoulders were added to construct a third computer model. Rather than creating the entire torso for the computer model, the shoulders were approximated as an oblate spheroid, where the height and depth of the shoulders had equal axis lengths, while the width of the shoulders had a longer axis length. Figure 2.26 shows this third computer model, with the head boundary in blue, microphone *a* in green and the microphone *b* in red. This third computer model produced the results shown in Figures 2.27 and 2.28.

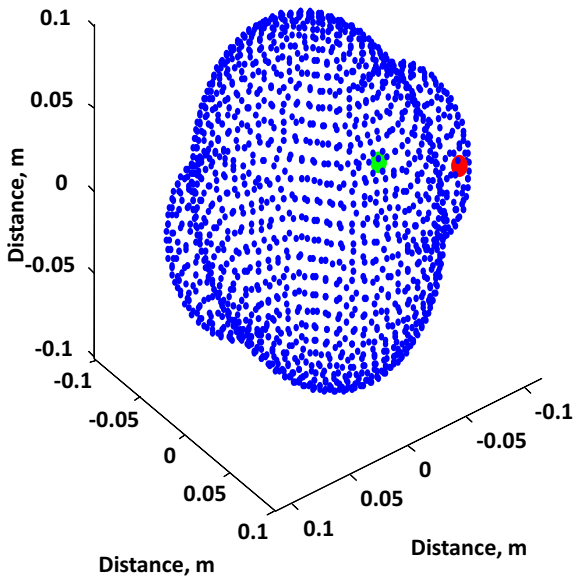


Figure 2.23 Ellipsoidal approximation for position of headphone microphones a (green) and b (red).

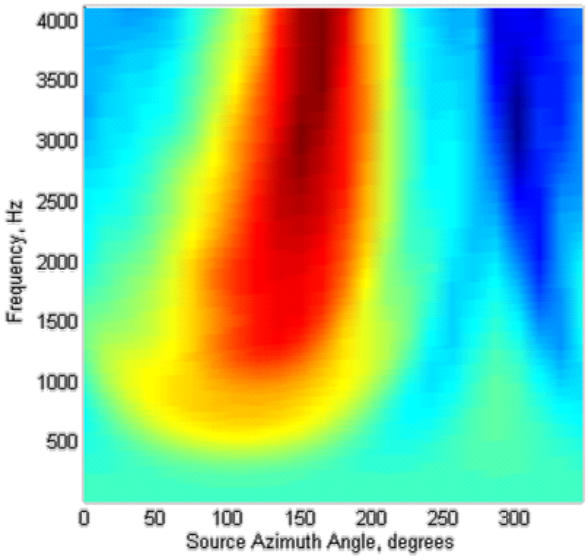


Figure 2.24 Ellipsoidal approximation results for microphone a.

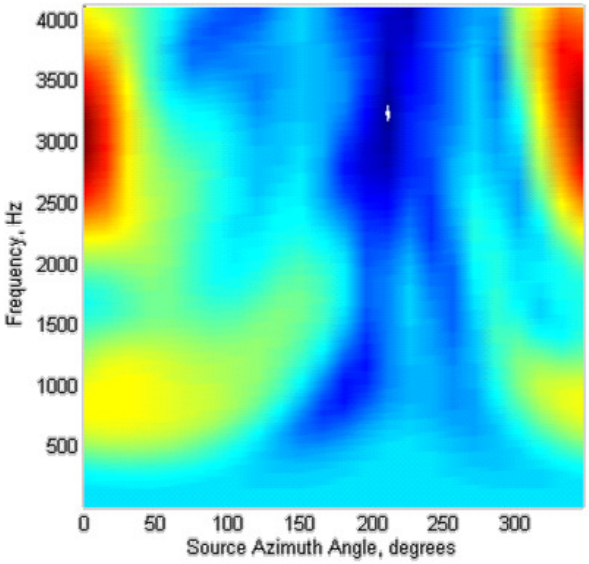


Figure 2.25 Ellipsoidal approximation results for microphone b.

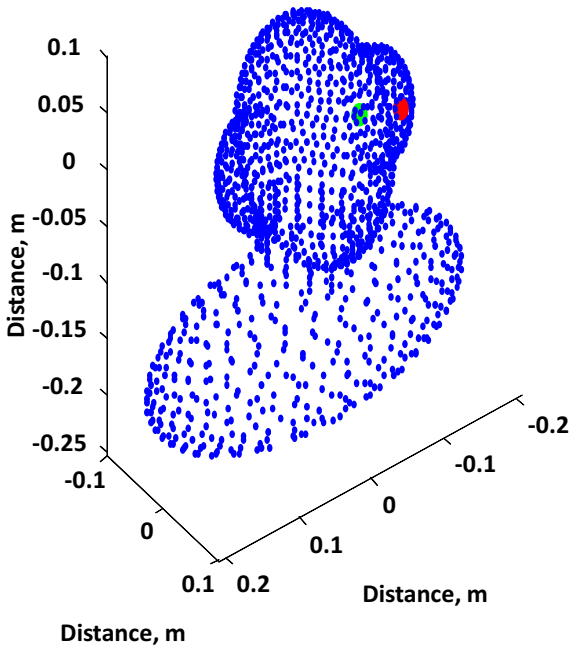


Figure 2.26 Ellipsoidal approximation with neck and torso for position of headphone microphones a (green) and b (red).

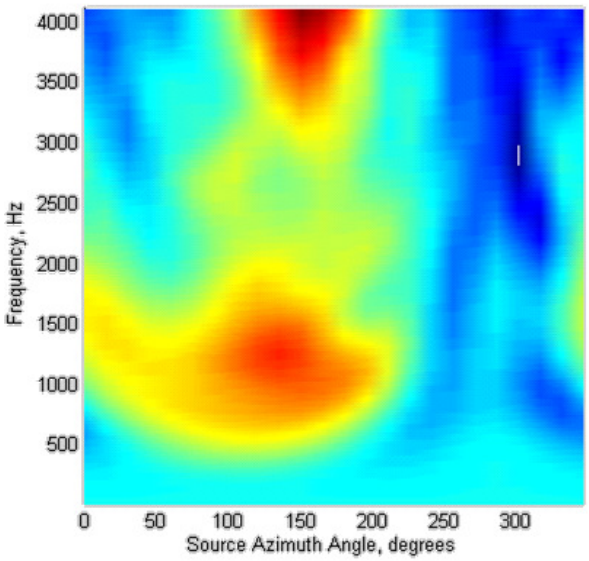


Figure 2.27 Ellipsoidal approximation with neck and torso results for microphone *a*.

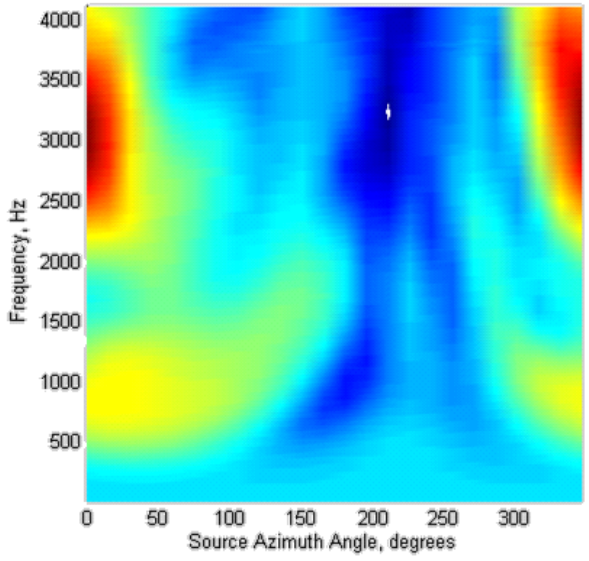
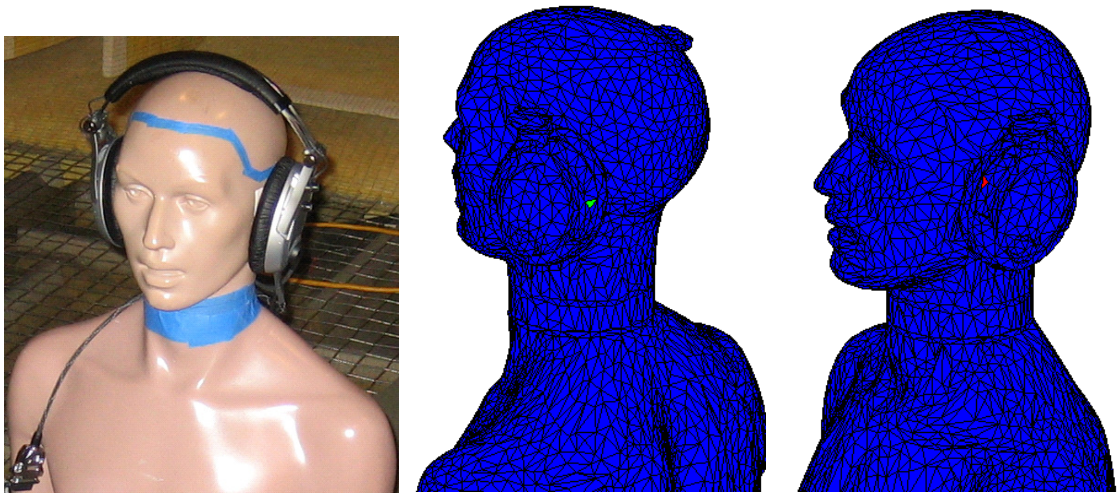


Figure 2.28 Ellipsoidal approximation with neck and torso results for microphone *b*.

2.4.8 Digital Scan of KEMAR

Creation of an accurate boundary element model of the human head allows performance assessments of prospective array designs without having to construct physical prototypes. Manual construction of a model using simple shapes such as spheroids and cylinders would be time consuming and cost prohibitive, so either a simpler approximation or method must be used to accurately the head. Also, a physical model of the virtual model should exist to allow experimental verification of the numerical results. The Knowles Electronics Manikin for Acoustic Research (KEMAR) manikin is generally recognized as the standard in replicating acoustic fields around the human head for experiments. With access to a 3-dimensional laser scanning machine, a high resolution computer model of a KEMAR head was created for numerical modeling. Another scan, shown in Figure 2.29 with the physical model, was performed with the KEMAR head wearing a set of headphones to provide a model for optimization of headphone microphone arrays. The results from this model are shown in Figures 2.30 and 2.31.



(a) KEMAR and Headset Array (b) Microphone a (c) Microphone b

Figure 2.29 KEMAR with headset array and 3-dimensional digital model of KEMAR and headset array showing positions of microphones a (green) and b (red).

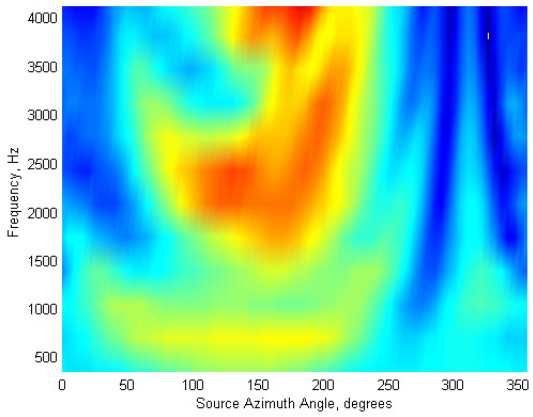


Figure 2.30 KEMAR digital scan results for microphone *a*.

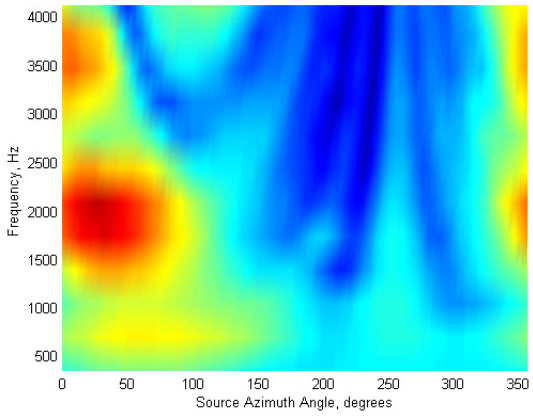


Figure 2.31 KEMAR digital scan results for microphone *b*.

CHAPTER 3

Experimental Validation of Array Manifolds

In addition to analytical and numerical methods, the array manifold matrix can be determined experimentally. This chapter presents a method for measuring array manifold vectors using an anechoic environment. Validation of the measurement method is shown through experimental results of array geometries presented in Chapter 2. Arrays are given in order of increasing geometric complexity from a circular array to a headset mounted array on a KEMAR manikin. Finally, array responses from the computational model of the KEMAR and headset array are presented and compared to experimental results to show a high level of agreement.

3.1 Experiment Setup

Once a microphone array has been constructed the array manifold can be experimentally measured within an anechoic environment to verify its performance. An anechoic, or sufficiently acoustically damped, environment is necessary to prevent reverberations from corrupting array manifold measurements. The process, referred to as array characterization, involves determining the sensitivities (adjustable gains, frequency response, spatial sensitivity, acquisition system sensitivities, etc.) of all of the elements within the system. The first step of array characterization is performed by playing a known sound stimulus through a reference speaker and measuring the resulting microphone responses. The microphone responses are then compared to the speaker stimulus signal to calculate

a transfer function relating sound from the speaker direction (assumed to be in the far-field) with the array response, the array manifold vector for that direction. To build up a complete array manifold matrix the process is repeated for all of the desired azimuth and elevation angles (and ranges if near-field responses are applicable).

Acquisition of the speaker signal can be performed in a number of ways, each of which will arrive at an acceptable measurement of the array response. The signal sent to the sound card, the voltage sent to the amplifier, or a high quality reference microphone can all be used as reference signals. The point in the signal chain where the reference signal is gathered will affect the total measurement of the array manifold, but assuming the effects are constant for the experiment, they will not adversely affect the measurement. For example, if the signal sent to the sound card is used as a reference, the dynamics of the sound card, amplifier, and speaker will all be captured within the array manifold measurement. So long as the behavior of each of those systems is linear, their responses can be accounted for. Figure 3.1 shows a headphone array undergoing characterization on a KEMAR manikin while using a reference microphone to measure the speaker stimulus signal.

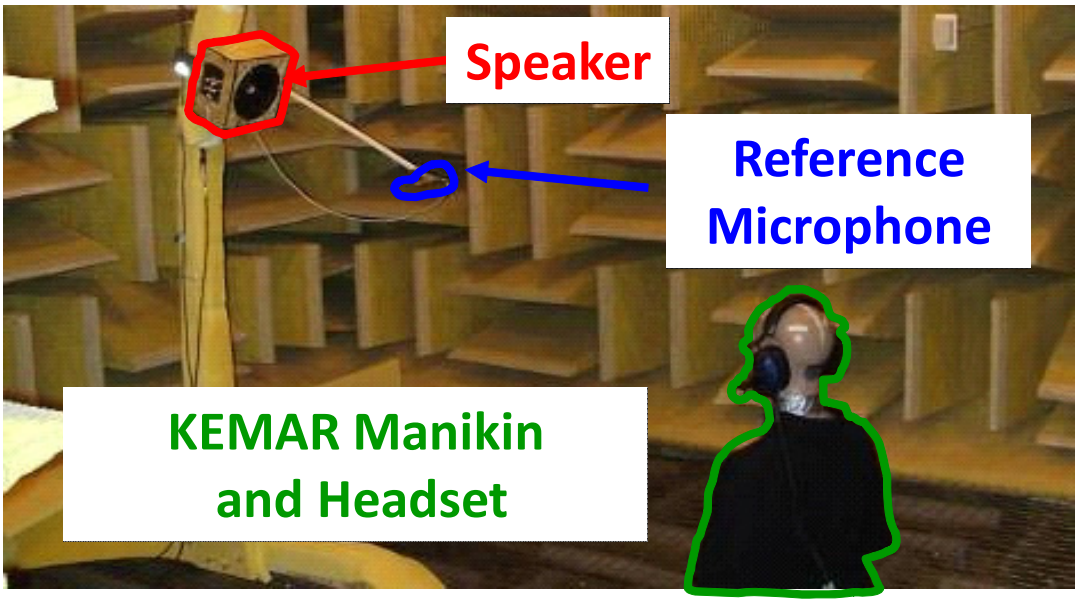


Figure 3.1 Characterization of a microphone array on a KEMAR manikin using a reference microphone.

The speaker signal selected as a reference stimulus also has implications on the characterization process. Single tones can be played if only narrowband characterization is desired, but there are many choices available for a broadband characterization. Rather than making separate measurements for each frequency, the orthogonality of frequencies can be exploited by characterizing all frequencies with a single measurement. Several broadband signal types are available, each with its benefits and drawbacks, described in greater detail in Appendix B.

3.2 Circular

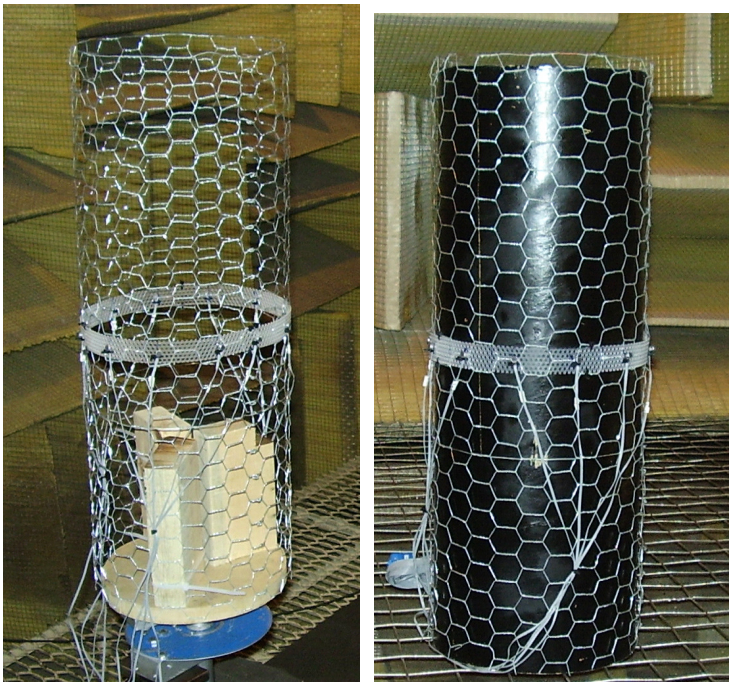
Figure 3.2 shows a free field and a diffracting circular array constructed for experimental data collection and analysis. The circular array was comprised of 12 microphones equally spaced around a 0.13m radius cylindrical wire mesh that was constructed to fit over a rigid cylinder. The purpose of the construction was to directly determine the effect of the absence or presence of a rigid cylinder on the response of the microphone array.

3.2.1 Free Field

The validation of the circular array microphone responses is presented for all azimuth angles and wavenumbers by examining the phase response of a single microphone from experimental measurements as compared to theoretical results. The magnitude response provides negligible insight due to nearly constant magnitude for all directions and frequencies. Figure 3.3 shows the microphone phase at all azimuth angles and wavenumbers. A few disparities exist but the free field circular array theory matches experimental measurements quite well.

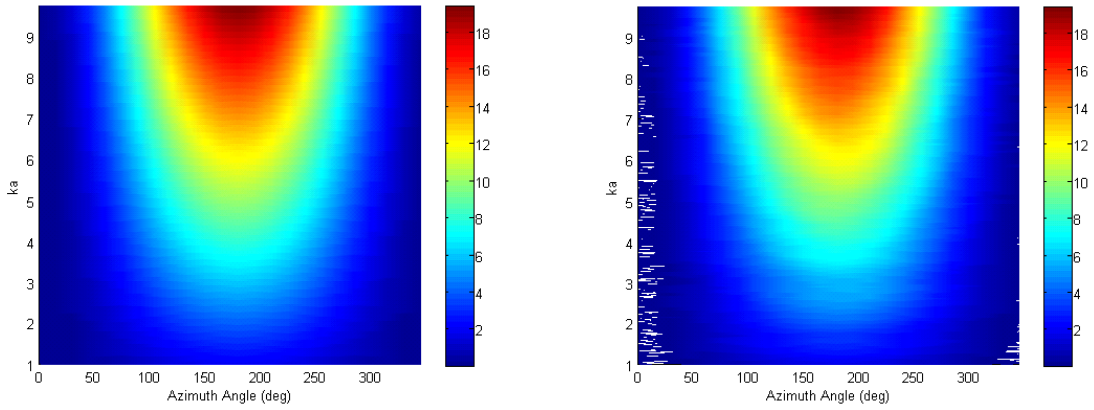
3.2.2 Diffracting

The diffracting circular array produces more distinct microphone responses than the free field case, as expected by theory. Figure 3.4 shows the phase response of a single



(a) Free Field (b) Diffracting

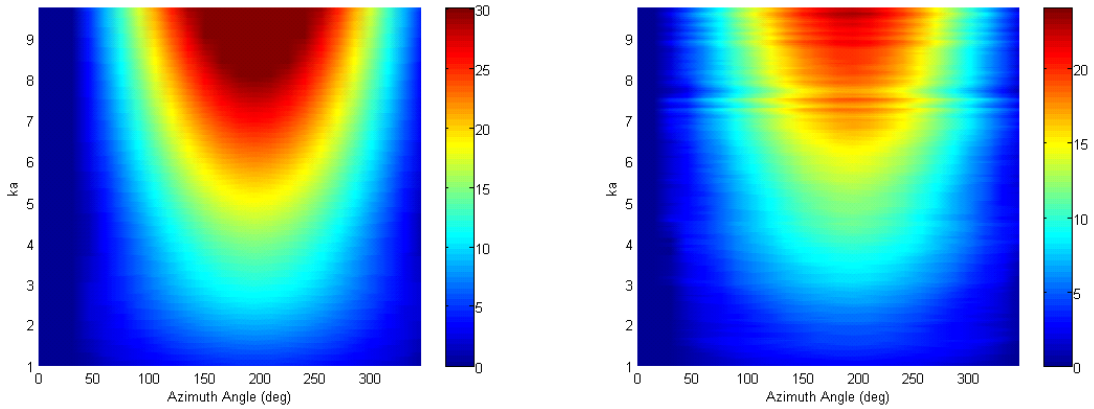
Figure 3.2 Circular arrays used for experimental data collection.



(a) Theoretical (b) Experimental

Figure 3.3 Phase at all azimuth angles and wavenumbers for a single microphone in a circular free field array of radius a=0.1m.

microphone at all azimuth angles and wavenumbers from theory (Figure 3.4(a)) and experiment (Figure 3.4(b)). In the diffracting case the magnitude differences at all azimuth angles are no longer negligible, as they are in the free field case. Figure 3.5 shows the magnitude response of a single microphone, located at an approximate azimuth angle of 200°, for all azimuth angles and wavenumbers. Note the magnitude rise when the source is pointing directly at the microphone (200°) and the magnitude shadowing when the microphone is on the far side of the cylinder (20°). At lower values of ka there is actually some constructive interference on the far side of the cylinder (20°) that leads to a magnitude response increase, but this behavior tapers off as ka increases.

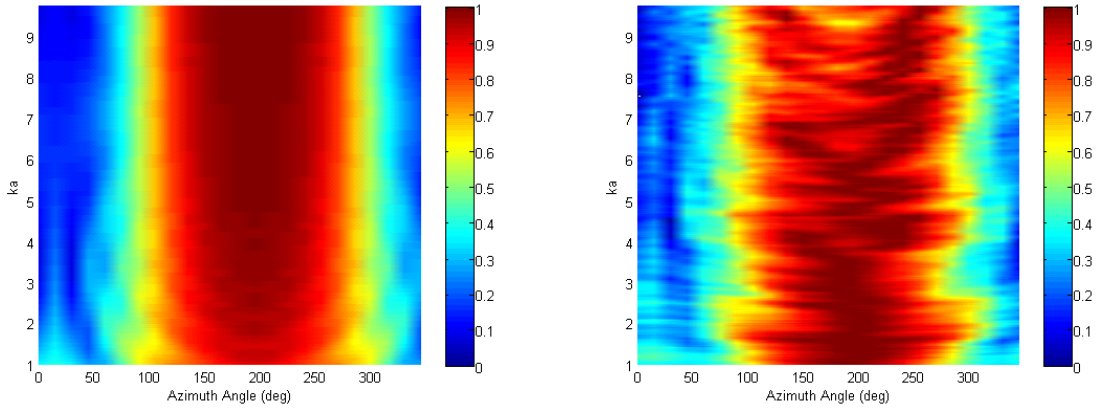


(a) Theoretical (b) Experimental

Figure 3.4 Phase at all azimuth angles and wavenumbers for a single microphone in a circular diffracting array of radius a=0.13m.

3.3 Spherical

Extending the design of a microphone array into 3-dimensions to represent the shape of a human head begins with a spherical microphone array. The symmetry of spherical arrays is useful for illustrating concepts, and more complex 3-dimensional geometries will be discussed in later sections. As in the circular array case, the spherical array can exist in the free field or on the surface of a rigid body, in this case a rigid sphere.



(a) Theoretical

(b) Experimental

Figure 3.5 Magnitude at all azimuth angles and wavenumbers for a single microphone in a circular diffracting array of radius a=0.13m.

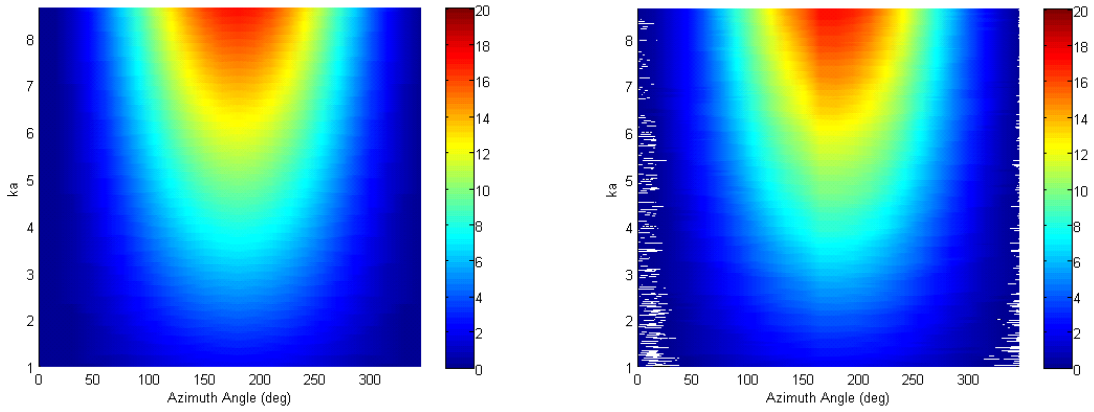
A great amount of research is currently being done into the applications of both free field and diffracting spherical arrays [36][37][38][39] as well as circular arrays mounted on diffracting spheres [40]. Two examples of spherical diffracting arrays, constructed for experimental data collection, are shown in Figure 3.6.

While a sphere is not as geometrically complex as the human head, a rigid sphere model of the head is a common analysis tool for researching ILD, ITD, and spectral cues [41][42], as well as a simple geometry for the initial testing of concepts and designs [43][44]. More complex models have been used for determining physical features responsible for spectral cues [35][45][46].

3.3.1 Free Field

The free field spherical array results are extremely similar to the free field circular array results due to the lack of diffractive effects. Figure 3.7 shows the microphone magnitude at all azimuth angles and wavenumbers. The free field spherical array theory matches experimental measurements quite well.

Figure 3.6 Spherical diffracting arrays used for experimental data collection and system testing.



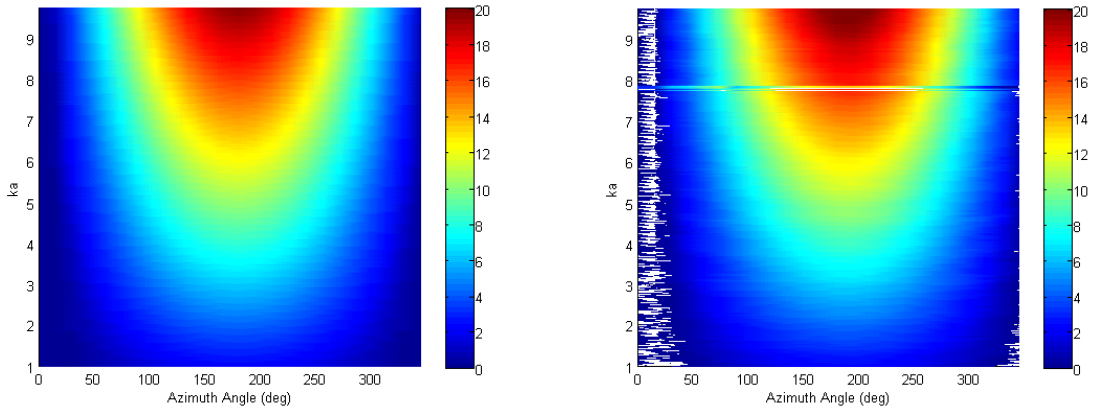
(a) Theoretical

(b) Experimental

Figure 3.7 Phase at all azimuth angles and wavenumbers for a single microphone in a spherical free field array of radius a=0.13m.

3.3.2 Diffracting

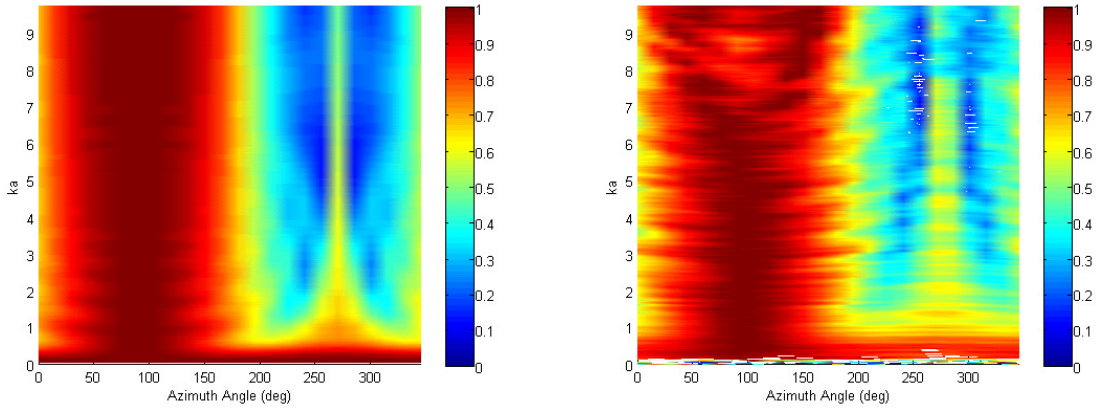
The comparison of theoretical and experimental results for a spherical diffracting array assumes a microphone mounted on the equator of a sphere at an azimuth angle of 90° with sources in the azimuth plane only. Figure 3.8 shows the phase response of a single microphone at all azimuth angles and wavenumbers from theory (Figure 3.8(a)) and experiment (Figure 3.8(b)). Figure 3.9 shows the microphone magnitude at all azimuth angles and wavenumbers. The agreement between theory and experiment is quite good for the diffracting spherical array and there are some interesting features to note that are caused by the diffraction. The magnitude increase seen in the diffracting circular array case is also present here (when the source is at 90°), but the shadowing of the microphone when it is on the far side of the sphere from the source (270°)has a different appearance. Constructive interference on the far side of the sphere appears to cause a magnitude rise for nearly all values of ka, compared to only lower values of ka for the diffracting circular array case.



(a) Theoretical

(b) Experimental

Figure 3.8 Phase at all azimuth angles and wavenumbers for a single microphone in a spherical diffracting array of radius a=0.13m.



(a) Theoretical

(b) Experimental

Figure 3.9 Magnitude at all azimuth angles and wavenumbers for a single microphone in a spherical diffracting array of radius a=0.13m.

3.4 Headset Approximation Models

While a spherical geometry provides an example of the array response for a 3-dimensional array, its approximation of a headset array on a human head is lacking. The diffractive effects from facial structures (i.e. the nose, eye sockets, etc.) and body geometry (i.e. the neck, shoulders, etc.) inevitably will change the response of an array mounted on the human head. Understanding the contribution to diffraction that each physical feature creates can be analyzed through gradual increases in the complexity of computational models. The effects of each model are described here for the models presented in Section 2.4.

3.4.1 Spherical Headset Model

This spherical approximation places microphones on the surface of a diffracting sphere. The directional responses of all microphones is expected to be nearly identical, as the microphones are closely located and no other structures (headphones, neck, etc.) exist to influence the responses. Since all of the microphones are located near where the ears would be, the expected maximum response for 2 microphones (see Figures 2.21 and 2.22) should appear when the left ear is pointing directly at the sound source, or 90° azimuth

angle based on the convention used here. As expected, 90° is where the maximum response occurs, and the response at other angles is consistent with spherical diffraction. Needless to say, almost none of the features present in the experimental data are present in this spherical approximation.

3.4.2 Ellipsoidal Headset Model

Compared to the spherical approximation, this ellipsoid approximation for the head and headphones matches more of the features present within the experimental data. Starting with Figure 2.24, the general shape of the response has changed from the symmetrical spherical approximation to a pattern where the angle of the maximum magnitude shifts with higher frequency. One of the features missing in the approximation in Figure 2.24 is the notch in the response that occurs above 2750 Hz in the experimental data. Up to 1500 Hz the approximation and experimental data appear to match well, but some shadowing below 100° and above 1500 Hz appears to be missing from the approximation. In Figure 2.25 the ellipsoid approximation matches the general shape of the experimental data quite well, but the size and location of features in the response are not correct. For instance, in the experimental data, a maximum in the response occurs at 2000 Hz and 25° that spans about 1000 Hz, whereas in the approximation, the corresponding maximum occurs at about 3250 Hz and 10°, spanning 1500 Hz. These differences suggest that the overall dimensions of the approximation may be some scaled version of the experimental dimensions, although there are still physical features missing from the data, such as the neck, shoulders, and torso. Another possibility is that the microphones in the simulation might not be in exactly the correct position or may have some other mounting or frequency response characteristics that are not duplicated exactly as in the experiment.

3.4.3 Head, Neck, and Shoulders Headset Model

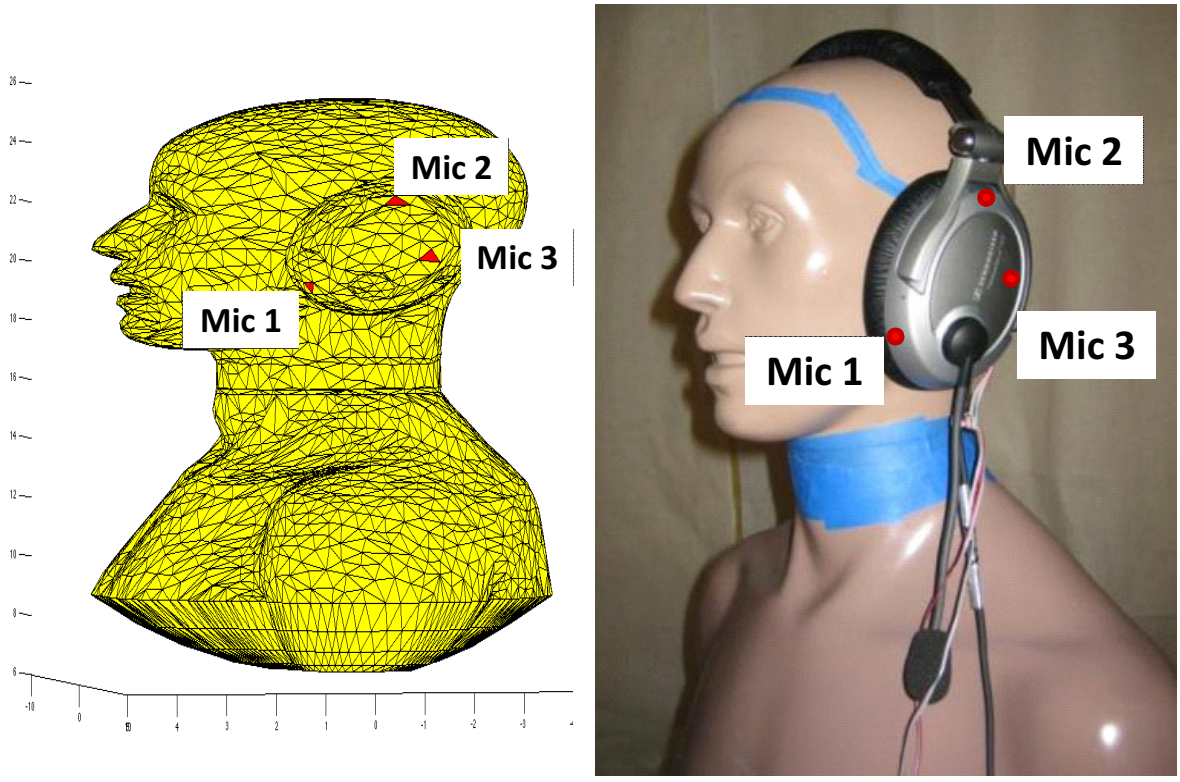
More of the features present in the experimental data have been resolved with the addition of the neck and shoulders, although an exact match for the experimental data

is still not present. The crescent shape in the response at 1000 Hz and 150° is better replicated in the simulation as compared to the experimental data, but now a break in the response between 2000 Hz and 3000 Hz has appeared that is not present in the experimental data. Also, the notch at frequencies higher than 3000 Hz has not been created, but part of that high frequency response has appeared, albeit shifted several degrees from its position in the experimental data. No change in the response in Figure 2.28 is expected, as microphone a sits on the headphone in the forward facing direction, ahead of the neck and shoulders that might cause some diffraction that would affect the response. Going along with this expectation, Figure 2.28 closely resembles Figure 2.25, where the neck and shoulders are not present.

3.4.4 Headset Mounted

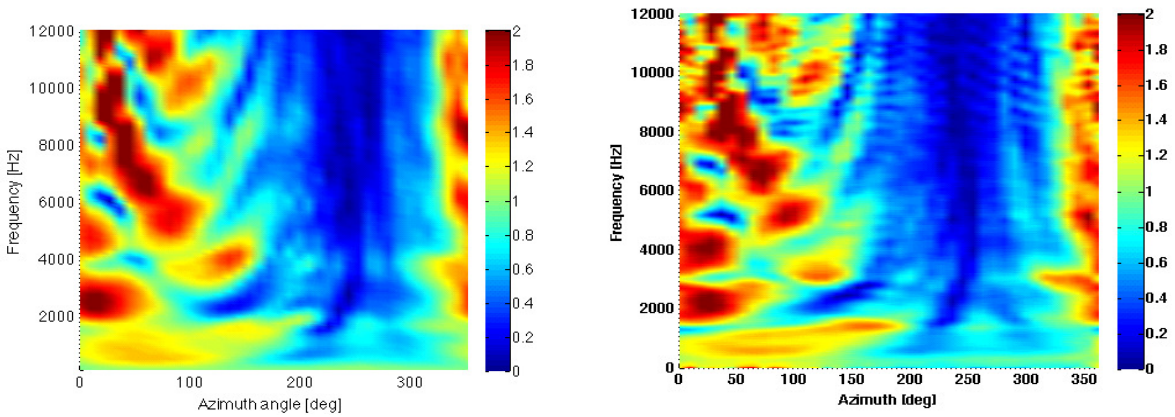
The examples given for circular and spherical arrays have only shown the response of a single microphone to sources in the azimuth plane. The symmetry in those cases lends itself to a simple description from only one microphone response. If another location around the cylinder or sphere were chosen, the magnitude plots previously presented (Figure 3.5 and Figure 3.9) would simply shift to the left or right to account for the change in position. When more features and details are added to the model of the head, this symmetry no longer applies. Placing microphones on a headset or helmet will have their responses inherently affected by the relative positions of diffracting structures such as shoulders, the nose, etc. Therefore, asserting that an acceptable agreement exists between theoretical and experimental models necessitates validation at multiple locations. Figure 3.10 shows the actual and numerical models of a KEMAR manikin wearing a headset equipped with 3 microphones. The responses of these microphones are compared from both numerical and experimental models in Figures 3.11, 3.12, and 3.13.

The agreement between the numerical approximation and the experimental results is superb. Even with the uncertainties in microphone placement in the experimental model to match the numerical model, the magnitude responses for all azimuth angles



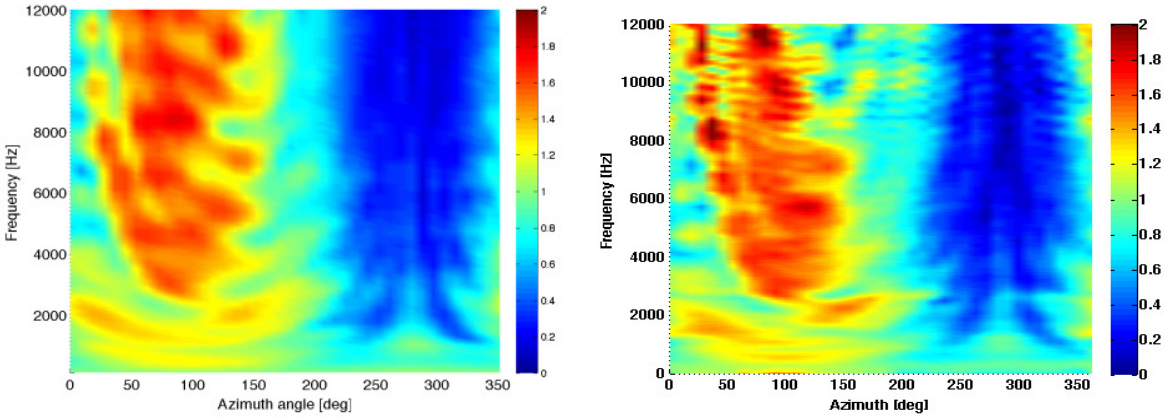
(a) Numerical model (b) KEMAR manikin

Figure 3.10 Real and digital models of KEMAR manikin used for validation of headset model with 3 microphone positions indicated.



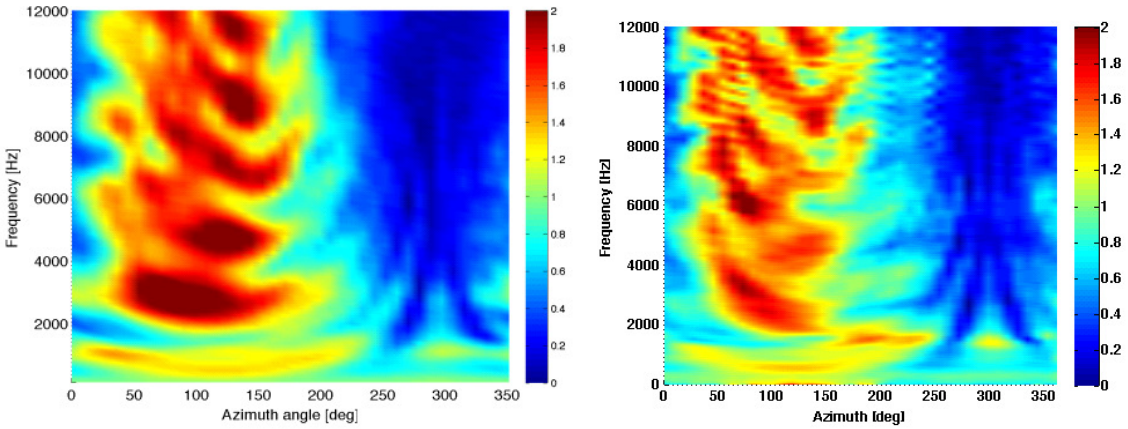
(a) Numerical (b) Experimental

Figure 3.11 Validation of numerical model with experimental results for microphone 1 from Figure 3.10.



(a) Numerical (b) Experimental

Figure 3.12 Validation of numerical model with experimental results for microphone 2 from Figure 3.10.



(a) Numerical (b) Experimental

Figure 3.13 Validation of numerical model with experimental results for microphone 3 from Figure 3.10.

and a frequency range up to 12 kHz show remarkable similarity. The numerical model shown in Figure 3.10 will be used to obtain results necessary for optimizing microphone positions in Chapter 7.

CHAPTER 4

Array Processing

Once an array manifold matrix is determined (whether analytically, computationally, or experimentally) it can be implemented in a processing strategy tailored to the particular array application. This chapter presents methods for calculating array filters that isolate sound coming from desired directions, localize sound sources in the environment, or minimize the influence of interfering sound signals. The first methods discuss how to sum all array signals, apply a fixed filter for any incoming sound, apply a dynamic filter that depends upon the characteristics of the incoming sound, or apply an adaptive filter that employs both fixed and dynamic filters at the same time. The last method presents a method to change the measured array signals to match the response of another array, illustrated with the example application of transparent hearing using a headset mounted microphone array.

Of the many goals and applications of arrays, most fall into two categories: beamforming and beampattern matching. Both categories are best explained by using the terms found in Equations 1.1 and 1.2 and Figure 1.3. Beamforming involves selecting array filters, $\mathbf{w}$, that preserve sound signals from desired angles while minimizing sound signals present in the selected noise model, $\mathbf{n}$. Achieving the best set of filters for a given application requires an accurate measurement or estimate of the array manifold vectors, $\mathbf{d}_\theta$, and noise model, $\mathbf{n}$. In beampattern matching the goal is to select array filters, $\mathbf{w}$, that recreate the response of a target array design. In general the target array design

is the human ears and the application is recreation of binaural sound fields, although surround sound recreation may implement the same techniques.

4.1 Beamforming

Microphone arrays are typically used to perform source localization, improve directive gain, or increase signal-to-noise ratio. To illustrate the improved performance afforded by using multiple microphones, a logical progression through different geometries will be used from a single microphone to a line array of microphones to a circular array of microphones to finally a spherical array of microphones. For the sake of simplicity in the presentation of concepts both time dependence and range dependence (i.e. signal decay with distance) will be neglected.

The process of using an array of sensors to sample a sound field (or electromagnetic field) and filter those samples to favor signals from a particular direction(s) is known as beamforming. Rather than relying on a physical structure that must be adjusted to receive sounds from one direction, thereby ignoring many or all other directions, an array of sensors can be electronically "steered" using signal processing methods to achieve the same purpose. The most recognizable examples of physical steering devices are satellite dishes for TV and parabolic microphone dishes for professional sports applications. A more general term for beamforming is spatial filtering; using an array to spatially sample a sound field at different locations and then combining those samples to receive sound from one direction while attenuating signals from other directions [47]. Beamforming strategies have been developed to suit a wide range of applications in both electromagnetic and acoustic modes, primarily starting in the respective areas of RADAR [48][49] and SONAR [50] (not surprisingly, both military applications). There are a multitude of beamforming techniques and algorithms, many tailored to specific niches, and here only those most suited to the head mounted microphone array application will be presented. There are two classes of beamformers that can be implemented: data independent and statistically optimum. Data independent beamforming schemes are normally found in

broad-band array designs where a frequency invariant beampattern is desired [51], such as in speech intelligibility applications. Statistically optimum beamforming is used when the statistics of the array signals are possibly changing over time and adaptive updating of beamforming filters is necessary to maintain peak performance.

4.1.1 Broadside Arrays

The simplest beamforming strategy to implement is one in which all microphone signals are added together, known as a broadside configuration. Referring to Equation 1.1 and Figure 1.3, a broadside configuration is the same as a set of filters, $\mathbf{w}$, where all entries are unity,

$$\mathbf{w} = [1 \quad 1 \quad \cdots \quad 1]^T \tag{4.1}$$

This terminology comes about because a source perpendicular to the axis of a line array (i.e. broadside to the array) is reinforced in the array output due to the summing strategy. That is to say that the maximum spatial sensitivity for this beamformer occurs when all microphones receive the same sound simultaneously. In contrast, if the microphone signals were delayed and added in such a way to reinforce a source collinear with the axis of the array, the array would be in an endfire configuration. Because it does not introduce any complicated processing methods, the broadside beamformer will be used to illustrate the benefits and drawbacks of array geometries from line to circular arrays. For the sake of completeness the single microphone case will be briefly discussed.

4.1.1.1 Single Microphone

The simplest case (and arguably not an array) is to have a single microphone. The entire performance of the microphone is based on its quality, operating range, and directionality. Assuming an ideal isotropic (omnidirectional) microphone with a suitable operating range, the measurable quantities are limited to the absolute magnitude and phase of signals measured. Estimation of bearing requires some differential time or magnitude measurement, neither of which is available without another microphone. Without

knowledge of the source signal's initial properties (power, frequency range, etc.), no range estimation is possible either. The signal to noise ratio of a single microphone can generally only be improved by increasing the signal strength, assuming the self-noise of the microphone is constant. Referring to Equation 2.3, neglecting decay and transient behavior of the sound source, and assuming no noise, the signal at microphone 1 due to the m^{th} source is

$$x_{1,m} = a_{1,m}e^{i\gamma_{1,m}} \tag{4.2}$$

where $a_{1,m}$ is the magnitude of the signal and $\gamma_{1,m}$ is the phase measured based on the propagation distance. As an aside, research has shown that specially shaped diffracting structures, referred to as neuromorphic, modeled after the human pinna can return some capability to perform monaural localization [52][53]. While such findings are promising and in line with the purpose of this dissertation, they are beyond the simple free-field assumption used thus far.

4.1.1.2 Line Array - 2 Microphones

Adding a microphone to the single microphone case yields a line array. The general form of the measured signals at each microphone, assuming no noise, are

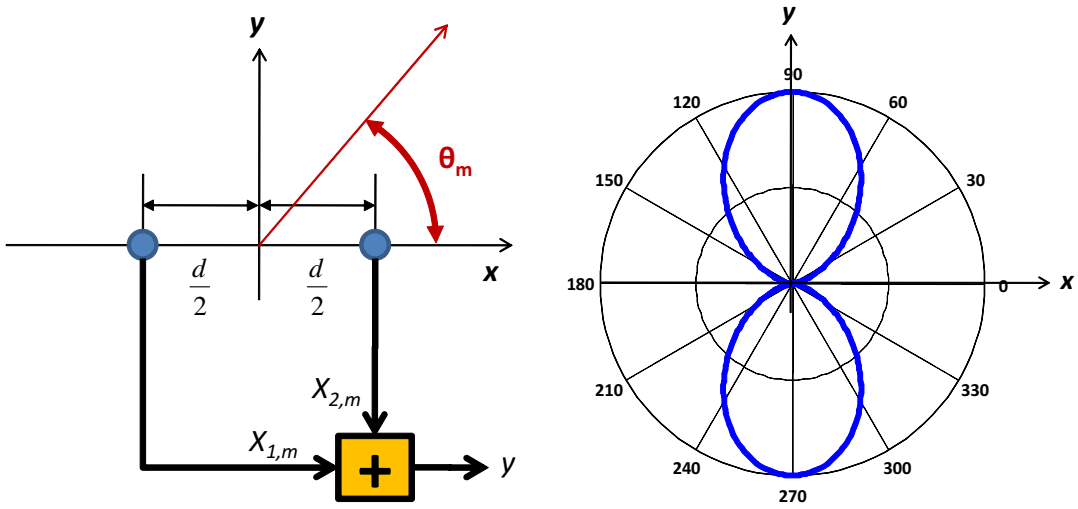
$$x_{1,m} = a_{1,m}e^{i\gamma_{1,m}} \tag{4.3}$$

$$x_{2,m} = a_{2,m}e^{i\gamma_{2,m}} \tag{4.4}$$

where $\gamma_{1,m} - \gamma_{2,m}$ represents the phase shift experienced by microphone 2 with respect to microphone 1.

For example, assume microphones 1 and 2 are placed at distances $-d/2$ and $d/2$ from the origin on the x-axis, as shown in Figure 4.1(a). The capabilities of this line array are based on the relative magnitudes and phases measured between the microphones. Consider a sound source with $k = \pi/d$ located on the positive x-axis at a distance r from the origin ($r >> d$). Since the wavelength of sound is exactly twice the spacing between the microphones, the measured signals will be exactly out of phase with one another (i.e. the microphone signals sum to zero). The same methodology can be repeated for far-field

sources at all angles in the $x-y$ plane to determine the spatial response, with the result shown in Figure 4.1(b).



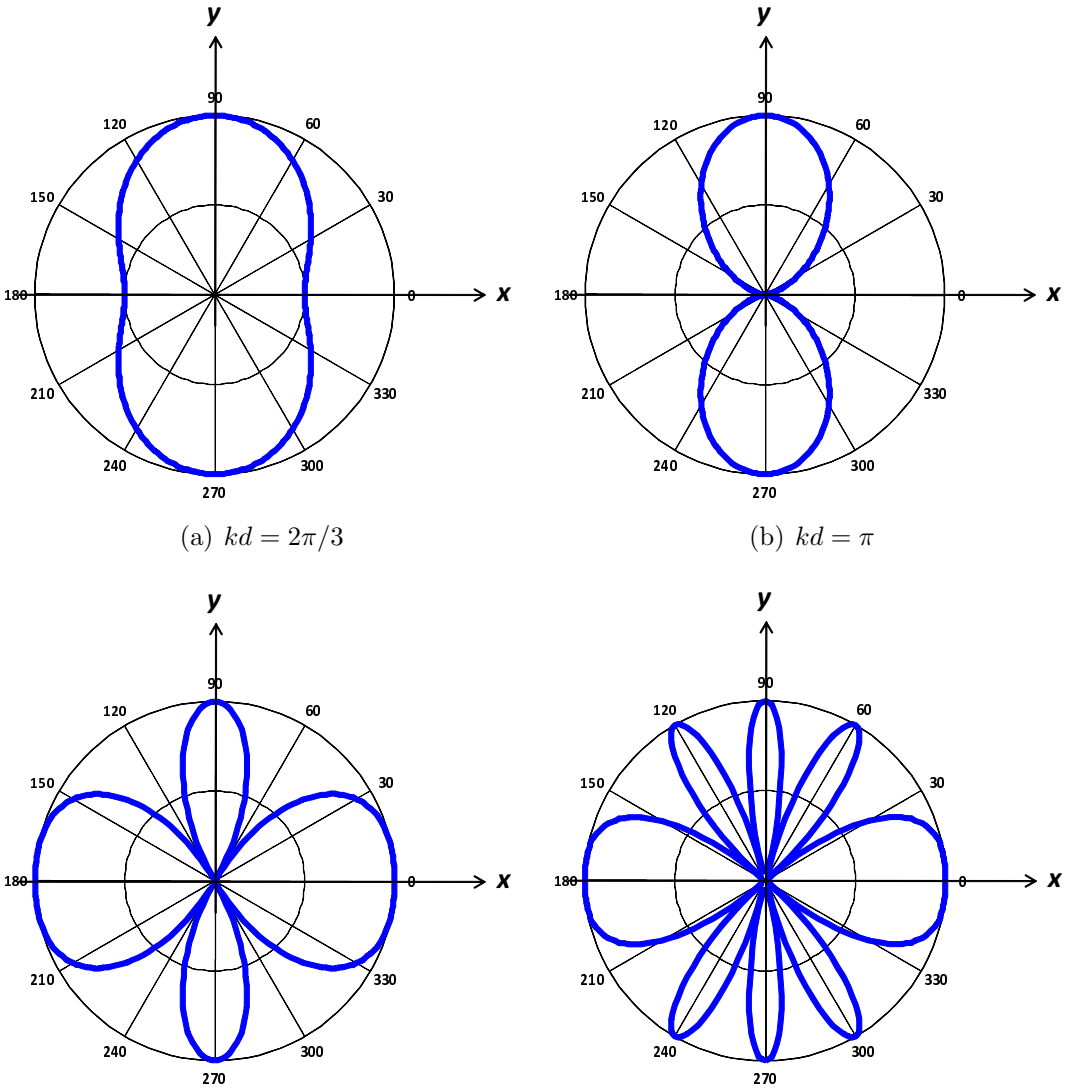
(a) 2 microphone array (b) Normalized broadside response

Figure 4.1 Summing strategy (broadside configuration) for a 2 microphone array with spacing d and $kd = \pi$.

The previous example deals with the broadside configuration spatial response for a single wavelength. Expanding the example to several other kd values ($2\pi/3$, π, 2π, and 4π) yields the broadside configuration spatial responses shown in Figure 4.2. The changing shape of the spatial response with increasing wavenumber is due to the relative phase between the microphones resulting in constructive and destructive interference patterns. Within each spatial response there can be several lobes, the one in the desired direction called the main lobe(s) and those of smaller magnitude called the side lobe(s). In the 2 microphone case the main and side lobes are of equal amplitude. To decrease the side lobe amplitudes more microphones need to be added to the array.

4.1.1.3 Line Array - Equal Spacing

As the number of microphones in an evenly spaced line array is increased beyond 2, a general form for finding the spatial response can be employed. The initial theory for equally spaced line arrays originates in electromagnetic antenna theory [54] and was



(a) $kd = 2\pi/3$ (b) $kd = \pi$

(c) $kd = 2\pi$ (d) $kd = 4\pi$

Figure 4.2 Normalized broadside configuration spatial responses for the 2 microphone line array shown in Figure 4.1(a) for different wavenumbers.

later extended to the acoustic domain [55]. For an array of N microphones on the x-axis centered on the origin with d inter element spacing, shown in Figure 4.3, the array response to a source at angle θ_m with the x-axis, assuming no noise, is

$$y = |a_{1,m}e^{i\gamma_{1,m}} + a_{2,m}e^{i(\gamma_{1,m}+\rho)} + \cdots + a_{n,m}e^{i(\gamma_{1,m}+n\rho)} + \cdots + a_{N,m}e^{i(\gamma_{1,m}+N\rho)}| \tag{4.5}$$

$$\rho = kd\cos\theta_m \tag{4.6}$$

Figure 4.4 shows broadside configuration spatial responses for $kd = \pi$ and arrays of 3, 4,

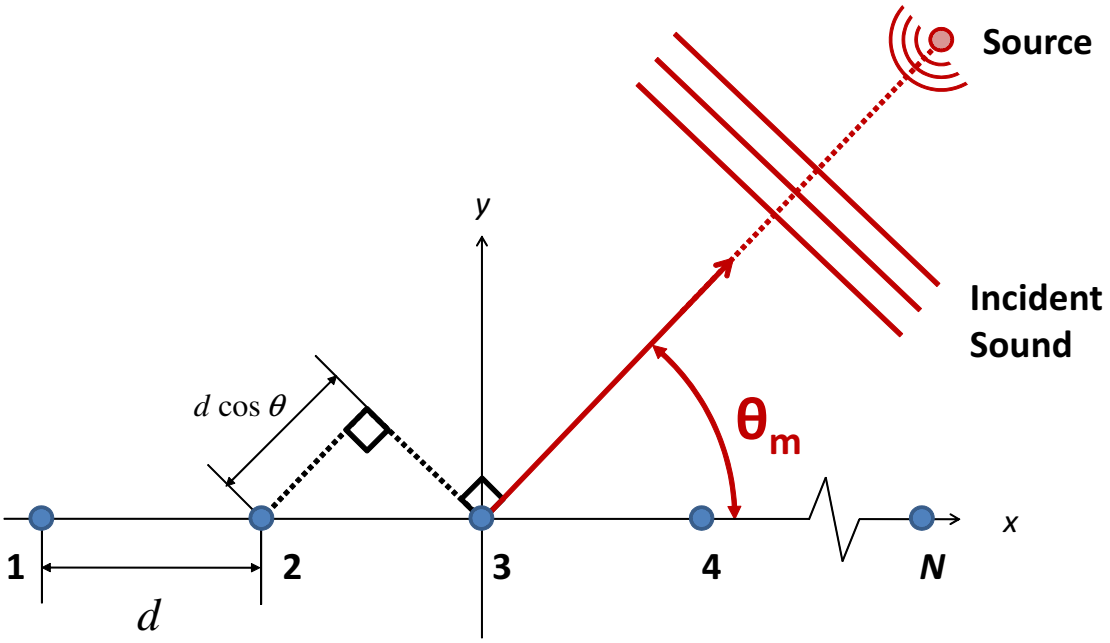
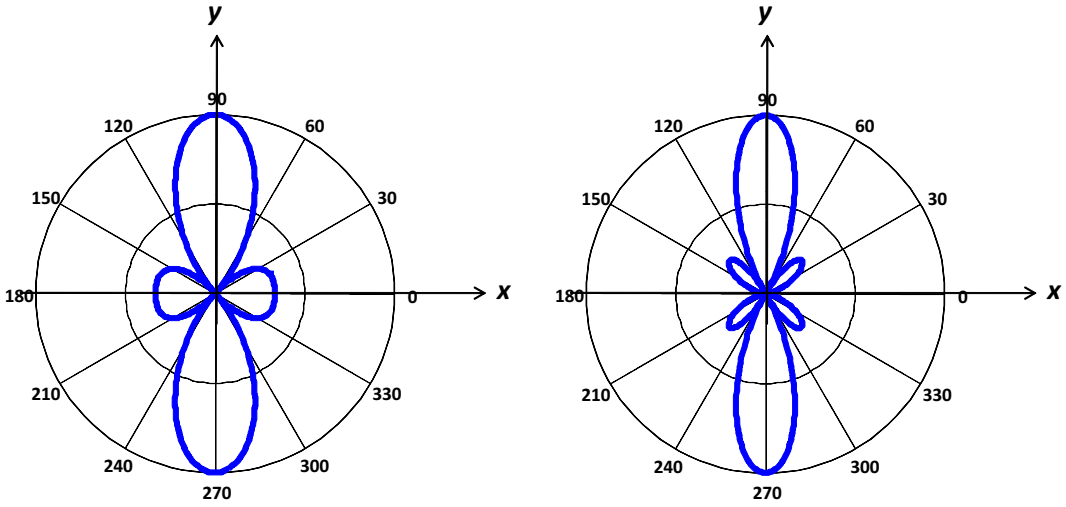


Figure 4.3 Line array with N elements and equal spacing d.

5, and 6 microphones with microphone spacing d and the array on the x-axis, centered on the origin.

The major shortcoming of a line array of microphones is the inability to determine whether sound sources are in front of or behind, above or below, the array. In this manner the line array acts much the same as duplex theory predicts human should behave; with a "cone of confusion [56]." The feature that shows this ambiguity is the symmetry of the spatial response across the axis of the line array. Extending the spatial responses

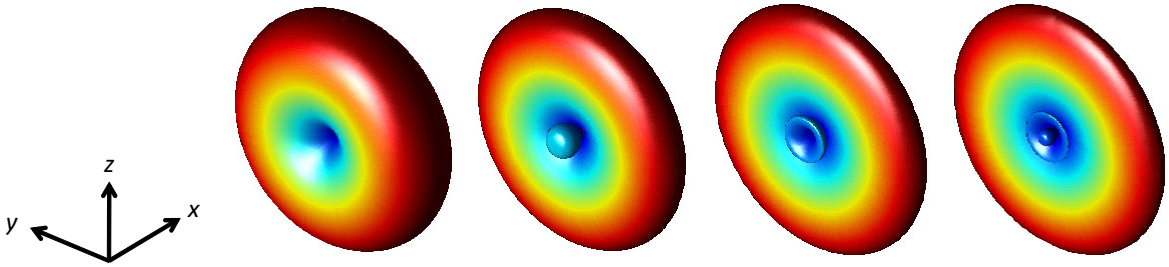


(a) $N = 3$ (b) $N = 4$

(c) $N = 5$ (d) $N = 6$

Figure 4.4 Normalized broadside configuration spatial responses for $kd = \pi$ and arrays with N elements with spacing d.

into the z-dimension shows the full nature of the ambiguity, shown in Figure 4.5. The feature to notice in Figure 4.5 is that the main lobe visible in the 2-dimensional spatial response has become donut shaped in the 3-dimensional representation. As microphones are added to the array the main lobe becomes narrower and more side lobes begin to form, but the overall ambiguity between sources in front of, behind, above, and below the array persists.



(a) $N = 3$ (b) $N = 4$ (c) $N = 5$ (d) $N = 6$

Figure 4.5 Normalized broadside configuration 3-dimensional spatial responses for $kd = \pi$ and arrays with N elements with spacing d.

4.1.1.4 Line Array - Unequal Spacing

While useful in presenting the underlying theory behind array processing, equally spaced line arrays are generally useful only when the source is narrowband in nature and present in the far-field [57]. As in the case of equally spaced line arrays, the theory behind unequally spaced line arrays originated in the field of electromagnetic antenna research [58] and was then adopted in acoustics research. As observed in Figure 4.2, the spatial response of a line array with equally spaced elements is highly sensitive to frequency. Unequal spacing between array elements is generally chosen to obtain a frequency invariant spatial response. For illustrative purposes an array design with logarithmic spacing is chosen using 15 microphones, shown in Figure 4.6. Logarithmic spacing is a popular design scheme for both antenna and acoustic line arrays, although prime number, arithmetic progression, elimination of multiples, and controlled cosine spacing strategies exist [59]. A thorough unequally spaced line array design procedure can be used to allow functional requirements to be translated into a physical layout [60].

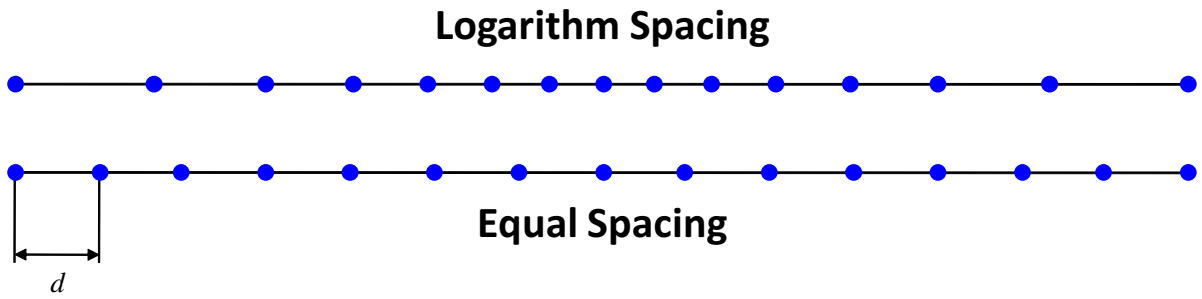
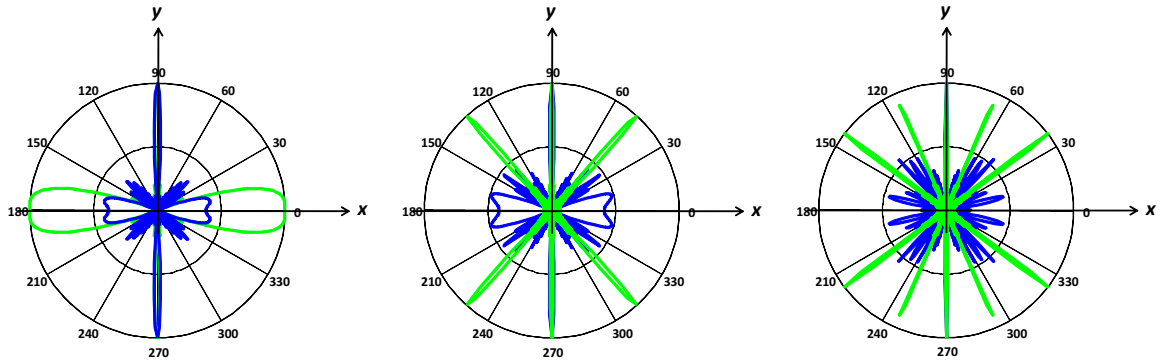


Figure 4.6 (top) Logarithm spacing of a 15 microphone line array and (bottom) equal spacing of a 15 microphone line array.

The benefit of using logarithm spacing in lieu of equal spacing is the relative size of the side lobes to the main lobe. Figure 4.7 shows the broadside configuration spatial responses for the 15 element microphone arrays given in Figure 4.6. The spatial responses are for $kd = 2\pi$ (Figure 4.7a), $kd = 3\pi$ (Figure 4.7b), and $kd = 5\pi$ (Figure 4.7c). In each case the side lobes of the logarithm spaced array are greatly reduced compared to the side lobes of the equal spaced array. The cost of side lobe reduction is a wider main lobe compared to the equal spacing array, but the overall performance of the logarithm array is much more constant over a wide band, or broad band, of wavelengths (i.e. frequencies). To further illustrate the point, Figure 4.8 shows the broadside configuration spatial responses for the logarithm and equal spacing configurations for a range of kd values from 2π to 8π. Within Figure 4.8, the desired main lobe is located at an azimuth angle of 90° and the side lobes are at surrounding azimuth angles. The drastic reduction in side lobe magnitude for the logarithmic array shows the performance benefit that can be gained from an unequal spacing strategy.

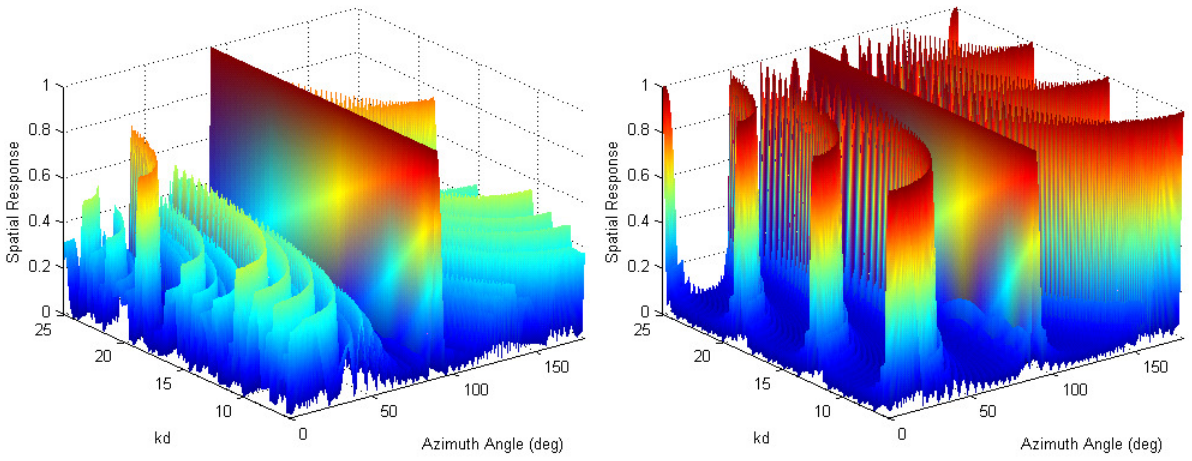
4.1.1.5 Circular Array

Expanding the array into 2-dimensions removes the "cone of confusion," although front-back ambiguity may still exist. There are many different 2-dimensional geometries used for various applications, notably equally spaced rectangular grids for auditoriums [61] and spiral arm arrays for aeroacoustic testing [62][63][64]. In an effort to



(a) $kd = 2\pi$ (b) $kd = 3\pi$ (c) $kd = 5\pi$

Figure 4.7 Normalized broadside configuration spatial responses for arrays with logarithm spacing (blue) and equal spacing (green) for values of kd.



(a) Logarithm (b) Equal

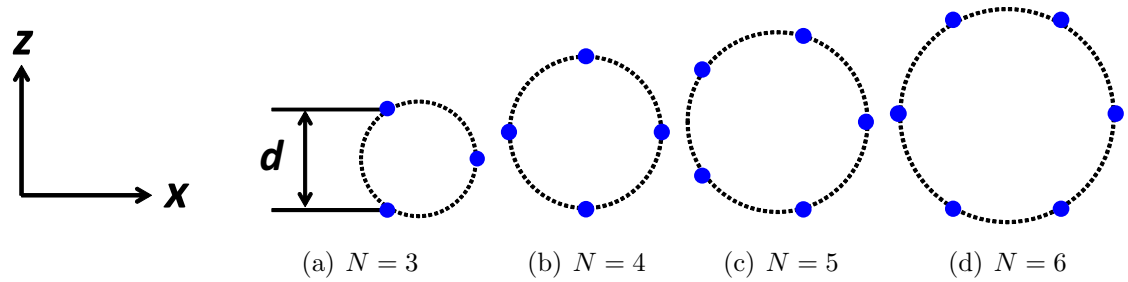
Figure 4.8 Normalized broadside configuration spatial responses for a range of $kd = 2\pi$ to $kd = 8\pi$ for different array spacings.

approach the complex shape of the human head, the 2-dimensional design discussed here will be a circular array with equally spaced elements around the circumference. As an extension of the examples given in the previous section, Figure 4.9 shows the geometries for 3, 4, 5, and 6 element circular arrays with element spacing d, where the position of the n^{th} microphone is

$$\theta_n = \frac{2n\pi}{N} \quad n = 1, 2, \ldots, N \tag{4.7}$$

$$x_n = \frac{d}{\sqrt{2 - 2\cos\frac{2\pi}{N}}} \cos\theta_n \tag{4.8}$$

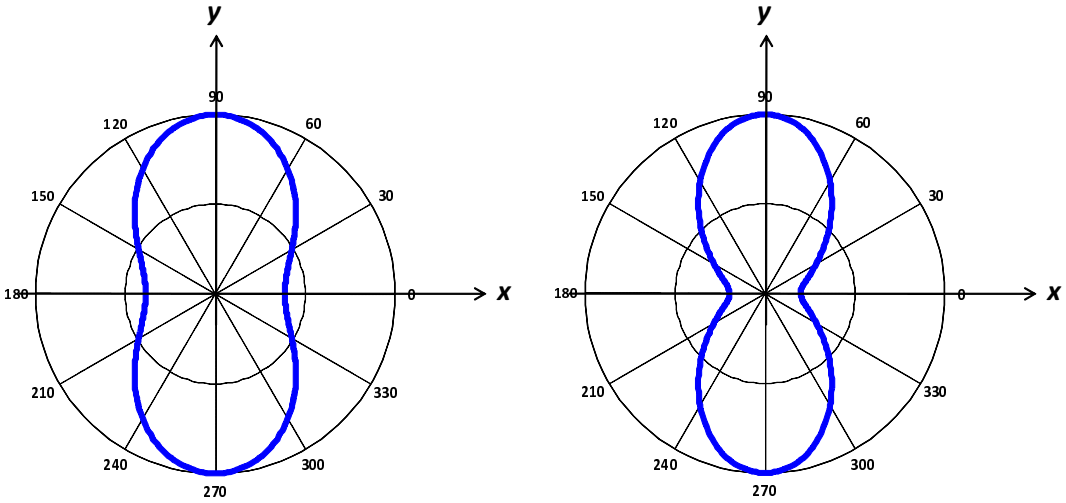
$$z_n = \frac{d}{\sqrt{2 - 2\cos\frac{2\pi}{N}}} \sin\theta_n \tag{4.9}$$



(a) $N = 3$ (b) $N = 4$ (c) $N = 5$ (d) $N = 6$

Figure 4.9 Geometries for circular microphone arrays with N equally spaced elements.

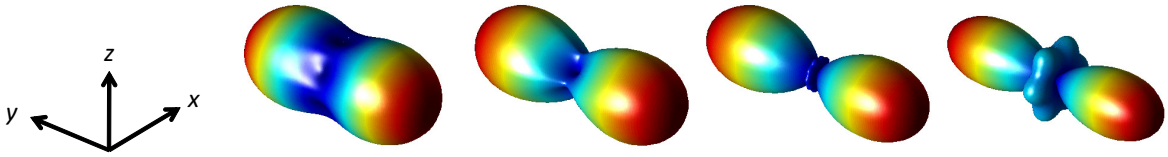
Employing the broadside configuration beamformer with a circular array will illustrate the benefits and drawbacks of arranging microphones in a symmetrical 2-dimensional pattern. For the purposes of maintaining the comparison between line and circular array results, the circular array will be considered coplanar with the $x - z$ plane. The spatial responses of the 4 circular arrays from Figure 4.9 are shown in Figure 4.10. The performance of these arrays would appear inferior to the equally and unequally spaced line arrays, but Figure 4.10 does not tell the whole story. While the azimuth spatial response of these arrays is not as good as their line array counterparts, the full 3-dimensional spatial response shown in Figure 4.11 illustrates the benefit of adding an extra dimension to the array design.



(a) $N = 3$ (b) $N = 4$

(c) $N = 5$ (d) $N = 6$

Figure 4.10 Normalized broadside configuration spatial responses for $kd = \pi$ for circular arrays shown in Figure 4.9.



(a) $N = 3$ (b) $N = 4$ (c) $N = 5$ (d) $N = 6$

Figure 4.11 Normalized broadside configuration 3-dimensional spatial responses for $kd = \pi$ for circular arrays shown in Figure 4.9.

The broadside configuration 3-dimensional spatial responses show that there is no "cone of confusion," although there is still front-back ambiguity for this array orientation. Another compromise that has been made in changing the geometry of the array is the effective frequency range, or aperture, of the array.

4.1.2 Data Independent - Delay and Sum (DSB)

The next level in complexity for a beamformer is the ability to "look" in different directions, accomplished through a delay-and-sum beamformer (DSB), generally synonymous with phased-arrays. Referring to Equation 2.3, for an array with N sensors, the array manifold vector for a source at angle θ_m is given by

$$\mathbf{d}_{\theta_m}(\omega) = [a_{1,m}e^{i\gamma_{1,m}} \quad a_{2,m}e^{i\gamma_{2,m}} \quad \dots \quad a_{n,m}e^{i\gamma_{n,m}} \quad \dots \quad a_{N,m}e^{i\gamma_{N,m}}]^T \tag{4.10}$$

where $\gamma_{n,m}$ and $a_{n,m}$ are the phase angle and response magnitude between the n^{th} sensor and the m^{th} source, respectively. The strategy behind the DSB is to compensate for the magnitudes and phases expressed within the array manifold vector to arrive at a set of constructively interfering signals that are added together. The resulting form of the DSB filter at a given frequency is

$$\mathbf{w}_{\theta_m} = \frac{1}{N}\mathbf{d}_{\theta_m} \tag{4.11}$$

Incorporating the DSB filter into the general filter model from Equation 1.1, repeated here,

$$\mathbf{y}(\omega) = \mathbf{w}^H\mathbf{x}(\omega) \tag{4.12}$$

yields

$$y = \frac{1}{N}\mathbf{d}^H_{\theta_m}\mathbf{x} \tag{4.13}$$

Whereas the broadside arrays discussed previously can only be employed for specific array-source orientations (perpendicular to the array axis, $\theta = 90°$), the DSB can be electronically "steered" to adjust for any array-source orientation. In effect, the DSB filter is created by inverting the phases in the array manifold vector, or compensating

for the phase differences of the received signals at each microphone. For example, if the array manifold vector is assumed to have the form of Equation 4.10, then the applied DSB filter will have the form,

$$\mathbf{w}_{\theta_m}^H(\omega) = [a_{1,m}e^{-i\gamma_{1,m}} \quad a_{2,m}e^{-i\gamma_{2,m}} \quad \dots \quad a_{n,m}e^{-i\gamma_{n,m}} \quad \dots \quad a_{N,m}e^{-i\gamma_{N,m}}] \tag{4.14}$$

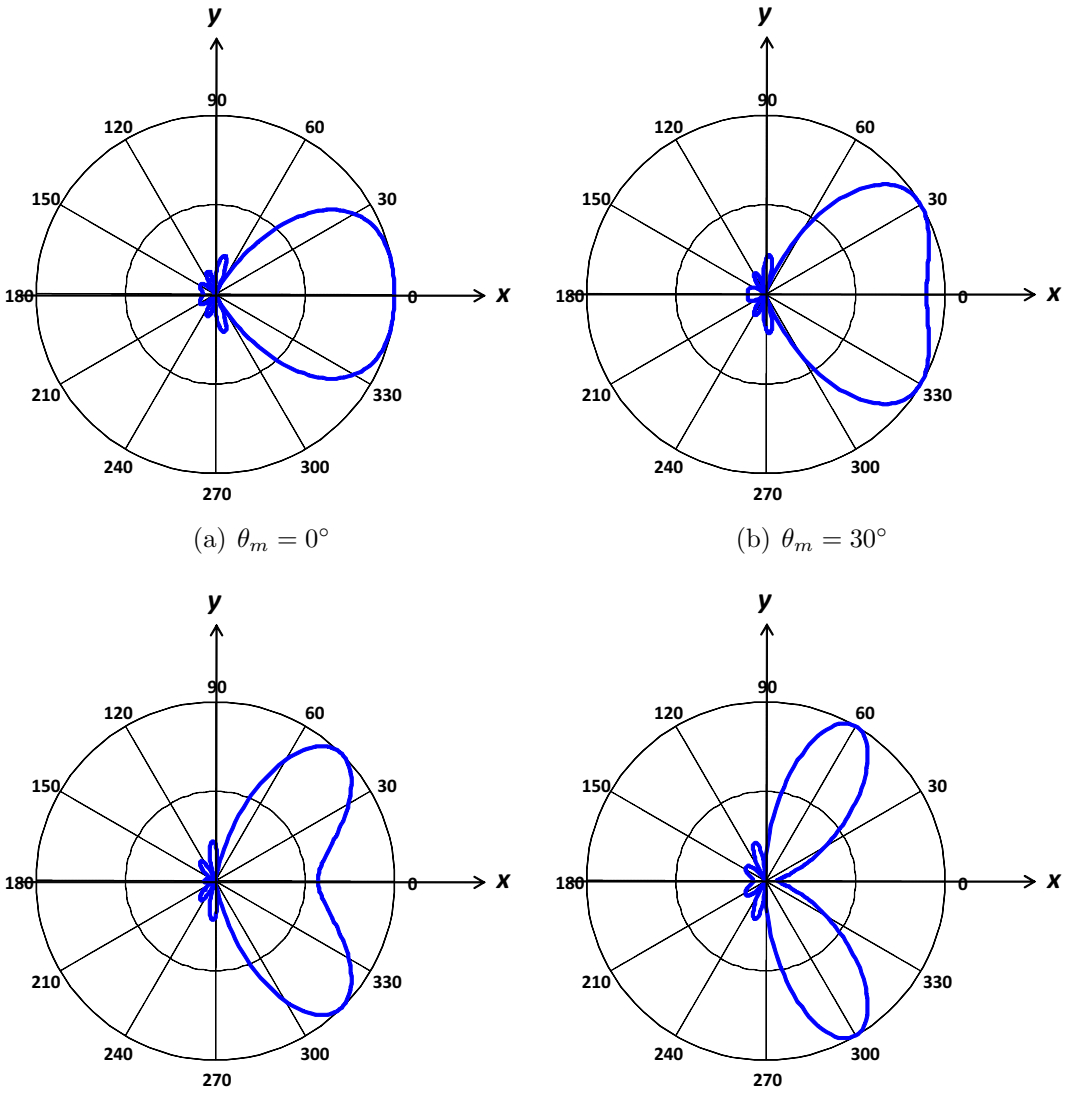
Strictly speaking, if the array is not in the free field, phase compensation is not exactly the same as delay-and-sum beamforming. However, for a free field array, delay compensation is equivalent to phase compensation. Using the equal spaced line array from Figure 4.6 as an example ($N = 15$), Figure 4.12 shows beampatterns for look directions of $\theta_m = 0°$, $30°$, $45°$, and $60°$ at $kd = \pi/4$.

4.1.3 Statistically Optimum - Superdirective(SDB)

While a conventional beamformer like the delay-and-sum strategy is relatively trivial to implement and completely independent of the sampled data, superdirective beamformers act to optimize the directivity (see section 5.2) of the array based on the statistics of the measured array signals. The optimization of the beamformer filters, $\mathbf{w}$, depends upon the measured noise, $\mathbf{n}$, and the array manifold vector, $\mathbf{d}$ (see Equations 1.1 and 1.2).

Superdirectivity, or supergain, was initially developed in the field of antenna arrays to address random errors present in sensors that led to poor performance of conventional beamformers [65]. These errors are a result of uncorrelated noise present at the sensors, generally thought of as self-noise or sensor noise. This is just one of many superdirective beamforming strategies that have been developed, a good historical review of which can be found in the literature [13]. Performance metrics, discussed in Chapter 5, can be used as constraints on the beamformer filter weights to achieve a desired array output [66].

In general a superdirective beamformer is designed to maintain the array response to a particular direction while placing constraints on other performance metrics. For example, maximizing array gain while constraining the sensitivity of the array to uncorrelated noise (known as the white noise gain). Popular superdirective strategies include minimum-



(a) $\theta_m = 0°$ (b) $\theta_m = 30°$

(c) $\theta_m = 45°$ (d) $\theta_m = 60°$

Figure 4.12 Normalized delay-and-sum beamformer beampatterns for 4 look directions at $kd = \pi/4$ for an equal spaced 15 microphone line array (from Figure 4.6).

variance distortionless response (MVDR) beamforming [14], where the array gain in the desired direction is constrained to unity while the overall beamformer output variance, or power, is minimized (hence the name), the previously mentioned MUSIC algorithm that uses an eigenvector approach to identifying source directions, and the enhanced minimum-variance (EMV) beamformer that can be adjusted to perform anywhere between the MVDR or MUSIC methods [67].

Obtaining the MVDR beamformer filters for a given frequency requires two constraints to be imposed upon the output signal. First, the signal response in the desired direction, also called the look direction, must remain undistorted, leading to the constraint

$$\mathbf{w}_\theta^H \mathbf{d}_\theta = 1 \tag{4.15}$$

which is to say that the MVDR filter, $\mathbf{w}_\theta^H$, when multiplied by the array manifold vector for the direction θ, $\mathbf{d}_\theta$, the magnitude of the result has unity magnitude and no change in phase (or distortion). The second constraint is a minimization of the overall output signal power

$$\min_{\mathbf{w}_\theta} \quad \mathbf{w}_\theta^H \mathbf{\Phi_{xx}} \mathbf{w}_\theta \tag{4.16}$$

where $\mathbf{\Phi_{xx}}$ is the cross spectral density matrix for the input signals $\mathbf{x}$

$$\mathbf{\Phi_{xx}} = E[\mathbf{x}^H \mathbf{x}] \tag{4.17}$$

where $\mathbf{x}$ contains the frequency components of the array response.

The derivation of the optimal filter, initially performed by Frost [68] using the method of Lagrange multipliers, results in

$$\mathbf{w}_\theta = \frac{\mathbf{\Phi_{xx}}^{-1} \mathbf{d}_\theta}{\mathbf{d}_\theta^H \mathbf{\Phi_{xx}}^{-1} \mathbf{d}_\theta} \tag{4.18}$$

At low frequencies where the array aperture appears small compared to a wavelength, $\mathbf{\Phi_{xx}}$ will be ill-conditioned, making the inversion sensitive to error in $\mathbf{d}_\theta$. A common remedy for ill-conditioning is to diagonally load $\mathbf{\Phi_{xx}}$ [69]

$$\mathbf{w}_\theta = \frac{[\mathbf{\Phi_{xx}} + \alpha \mathbf{I}]^{-1} \mathbf{d}_\theta}{\mathbf{d}_\theta^H [\mathbf{\Phi_{xx}} + \alpha \mathbf{I}]^{-1} \mathbf{d}_\theta} \tag{4.19}$$

where α is a small scalar relative to the norm of $\mathbf{\Phi_{xx}}$. This diagonal loading has essentially the same effect as adding uncorrelated noise to the sensor signals. In the limit of very large α the MVDR solution converges to the delay-and-sum beamformer. On the other hand, when α reaches the value of the noise covariance, $-\sigma_{\mathbf{n}}^2$, the response corresponding to the MUSIC algorithm is reached [70].

4.1.3.1 Diffuse Noise

Consider a sound field with M equally distributed sources in the far-field. For an array with N microphones the array manifold matrix, $\mathbf{D}$, will have dimensions N-by-M. Assuming a set of independent (i.e. uncorrelated), unit amplitude, far-field sources [39], where the cross spectral density matrix for that set of sources can be expressed as

$$\mathbf{\Phi}_{sources} = \mathbf{I} \tag{4.20}$$

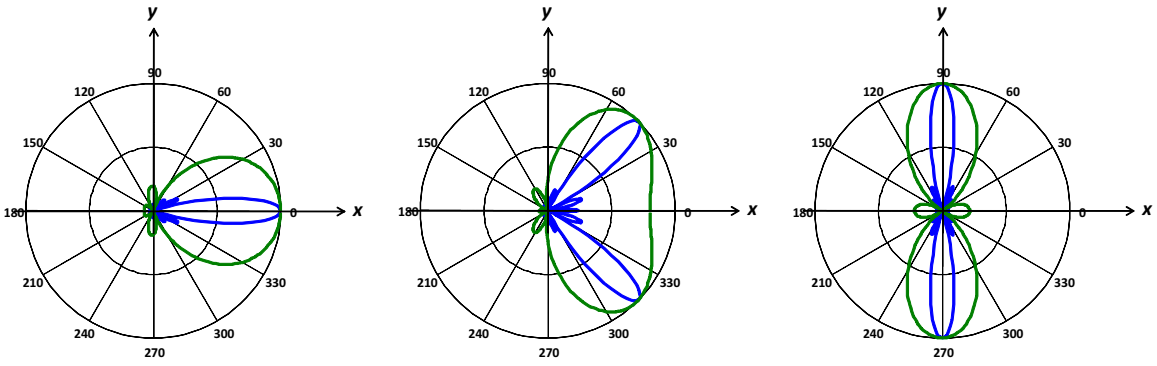
the cross spectral density matrix for the array signals, $\mathbf{\Phi_{xx}}$, can be calculated from the array manifold matrix via the relationship

$$\mathbf{\Phi_{xx}} = \mathbf{D}\mathbf{\Phi}_{sources}\mathbf{D}^H = \mathbf{D}\mathbf{D}^H \tag{4.21}$$

When the sources are equally distributed in the azimuth plane, the sound field is said to be cylindrically isotropic. Extending the equal distribution of sources into all 3-dimensions yields a spherically isotropic sound field. In the previous equation, only the source angles represented in the array manifold matrix will be incorporated into the cross spectral density matrix. When calculating the MVDR beamformer filter weights for a given look direction, the sensitivity to all other directions represented in the cross spectral density matrix will be minimized. This property can be exploited for applications where suppression of sound from specific directions is important.

If a measurement or estimate of the array manifold matrix is not available, an analytical formulation of the cross spectral density matrix may exist. For an array in the free field, $\mathbf{\Phi_{xx}}$ can be represented in terms of *sinc* functions [69]. For diffracting arrays, $\mathbf{\Phi_{xx}}$ must be written in terms of functions that describe the acoustical scattering produced by each mounting body [71].

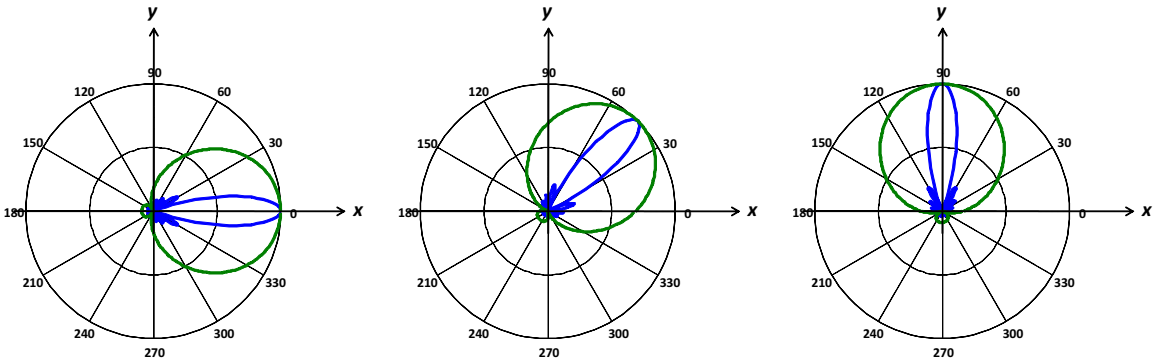
The performance disparity between conventional and superdirective beamformers is easily visualized through the array beampattern. Consider a 20 element line array with elements equally spaced apart distance d at a kd value of $\pi/10$. The beampatterns for conventional (blue) and MVDR (green) beamformers are shown in Figure 4.13 for look directions of (a) 0° (endfire), (b) 45°, and (c) 90° (broadside). For this example the sound field was assumed to be cylindrically isotropic, i.e. equal sound magnitude and distribution at all azimuth angles. Notice the drastic improvement in the beamwidth (discussed in Chapter 5) and the minimization of side lobes for the MVDR compared to the conventional beamformer. While an optimum beamformer has been shown to greatly improve directional gain, the ambiguity due to the geometry of the array persists.



(a) 0° (b) 45° (c) 90°

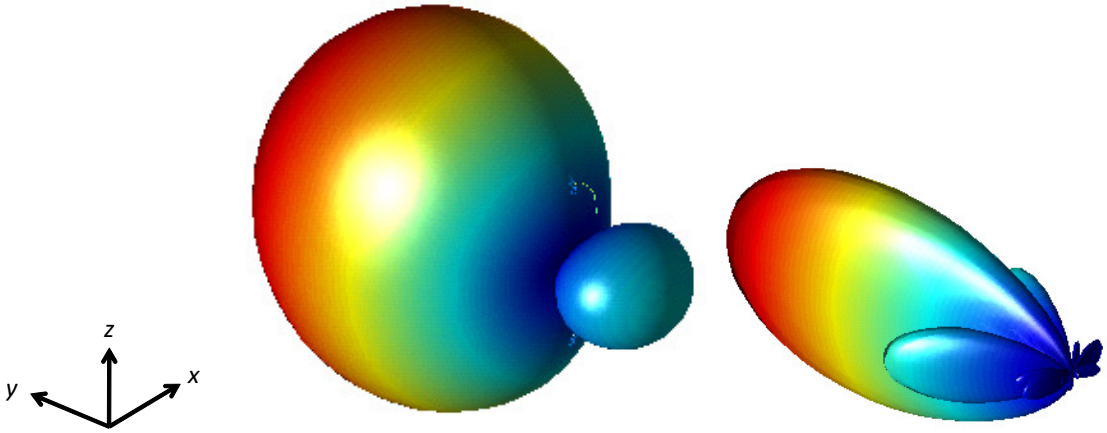
Figure 4.13 Normalized beampatterns for MVDR (blue) and DSB (green) beamformers for a 20 element equally spaced line array (spacing d) for a kd value of $\pi/10$.

Using the DSB and MVDR beamformers with a 20 element circular array with d spacing between elements and a kd value of $\pi/10$ yields the beampatterns shown in Figure 4.14 for look directions of 0°, 45°, and 90°. For Figure 4.14 the array is oriented parallel with the $x-y$ plane. Notice the lack of ambiguity when the look direction is changed and the consistency in both beampatterns for each direction. One further illustration of the difference between conventional and statistically optimum beamforming can be seen in the 3-dimensional beampatterns in Figure 4.15. The red portion of each beampattern indicates the highest gain (main lobes) while the blue portions indicates the areas of highest suppression.



(a) 0° (b) 45° (c) 90°

Figure 4.14 Normalized beampatterns for MVDR (blue) and DSB (green) beamformers for a 20 element equally spaced circular array (spacing d) for a kd value of $\pi/10$ and different arrival angles.



(a) Conventional (b) MVDR

Figure 4.15 Normalized 3-dimensional beampatterns for conventional and MVDR beamformers for a 20 element equally spaced circular array (spacing d) for a kd value of $\pi/10$.

4.1.4 Generalized Sidelobe Canceller(GSC)

Another adaptive technique for preserving signals of interest while minimizing the contributions from other signals is called the generalized sidelobe canceller (GSC). There are 2 main filtering paths in the GSC: a conventional beamformer in the direction of the sound and a set of adaptive beamformers in all directions except the desired direction. While an infinite number of blocking filters exist, the number of blocking filters implemented is limited to the number of degrees of freedom of the system, $N-1$. Attempting to implement more blocking filters than the degrees of freedom of the system will result in a filter compromise.

The result of this strategy is a beamformed signal in the look direction and a beamformed signal in all other directions, or a signal that mostly contains sound from the look direction and a signal that mostly contains sound from all other directions. An adaptive filtering method is then employed to remove any information that is correlated between the look direction signal and the non-look direction signal. The least-mean-squared (LMS) adaptive algorithm used is described in section 4.1.4.1. A simplified diagram of the GSC is shown in 4.16.

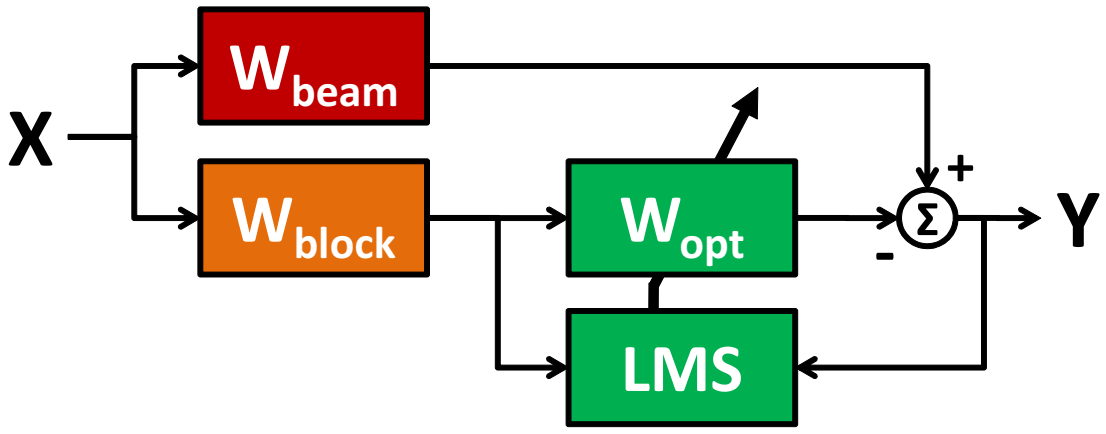
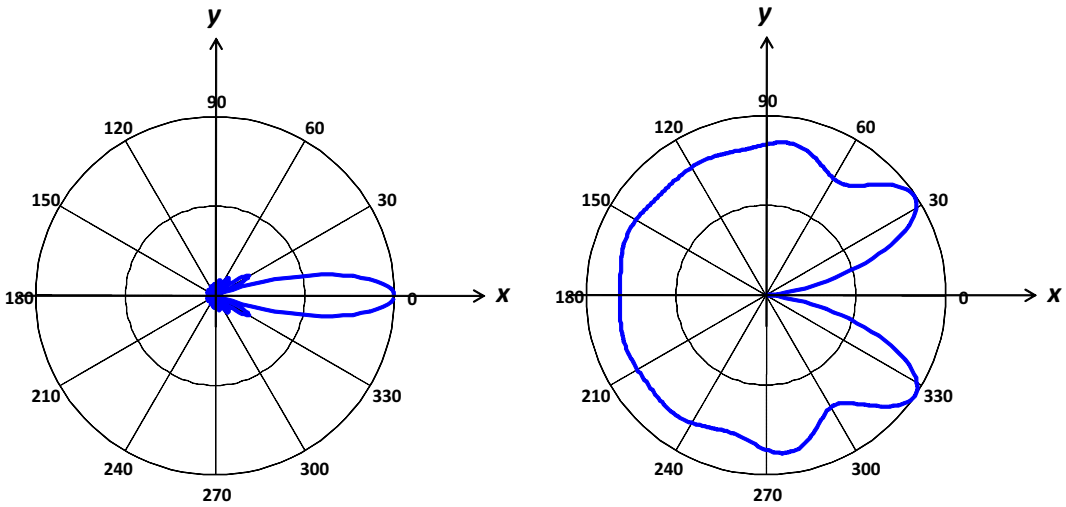


Figure 4.16 Generalized sidelobe canceller filtering block diagram.

Consider the 20 microphone circular array with equal element spacing d from section 4.1.3.1. For a desired sound source at an azimuth angle of 0°, the beam and block filter beampatterns are shown in Figures 4.17(a) and 4.17(b), respectively. Assuming a sta-

tionary interfering sound source at an azimuth angle of 52° and applying the GSC yields the beampattern shown in Figure 4.18, compared to the original beam only beampattern.



(a) Beam (b) Block

Figure 4.17 Beampatterns for "beam" and "block" filters for 20 microphone circular array with equal spacing d and $kd = \pi/10$.

Creation of the block filter is accomplished through a similar matrix inversion method to what was implemented in section 4.1.3 for the calculation of the MVDR filter. Rather than constraining the magnitude of the response in a particular direction to unity, this constraint for this filter is to minimize the response magnitude in a particular direction. However, if this response minimization in one direction were paired with the constraint of a minimum power output, the trivial solution would inevitably result. To avoid the trivial solution the calculation of the CSD matrix $\mathbf{\Phi_{xx}}$ will apply a very large weighting to the desired direction and unity weightings to all other directions. Heavily weighting the desired direction in the CSD matrix before inversion will result in a very small weighting in the final filter. The form of the weighted CSD matrix is

$$\mathbf{\Phi_{weighted}} = \mathbf{D}\Lambda\mathbf{D}^H \tag{4.22}$$

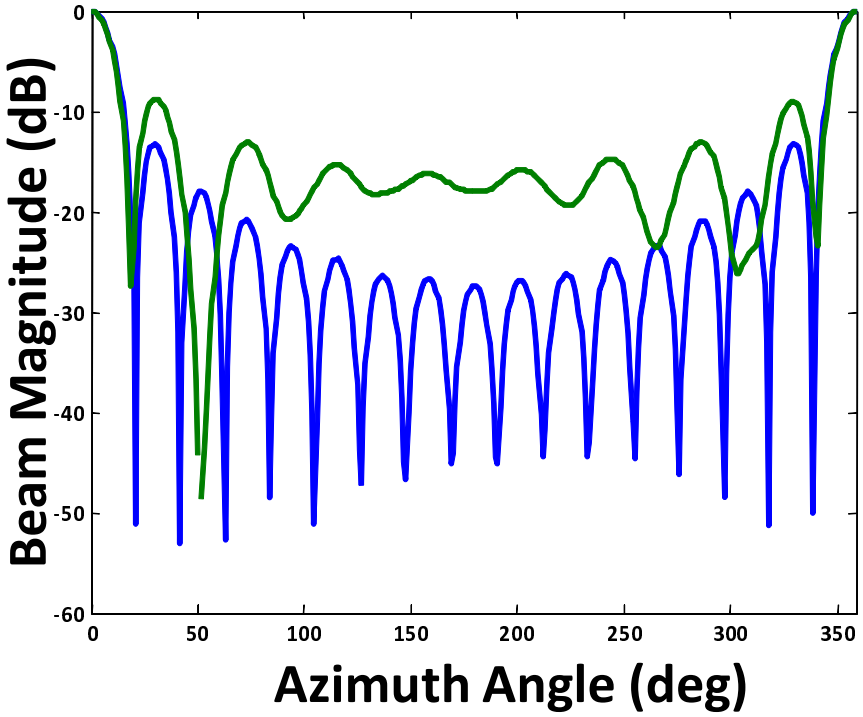


Figure 4.18 Adapted beampattern (green) after application of GSC method for 20 microphone circular array with equal spacing d, $kd = \pi/10$, and an interferer at 52°, compared to the "beam" filter beampattern (blue).

where the form of Λ is

$$\Lambda = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \tag{4.23}$$

where A_m is a very large constant compared to unity for the m^{th} direction, the direction whose response is to be minimized. The calculation of the block filter is then

$$\mathbf{w}_{block} = \mathbf{b}\mathbf{D}^H \mathbf{\Phi}^{-1}_{\mathbf{weighted}} \tag{4.24}$$

where $\mathbf{b}$ is a blocking vector

$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{4.25}$$

where the zero term is the m^{th} term in the vector corresponding to the m^{th} direction.

4.1.4.1 Least-Mean-Squared (LMS) Algorithm

Beginning in the form of an electronic circuit in 1959 [72], the least-mean-squared (LMS) adaptive algorithm is designed to converge to the optimum filter. The inputs to the LMS algorithm are a reference signal, $\mathbf{u}(k)$, and a disturbance signal, $\mathbf{d}(k)$. The LMS algorithm is used to adaptively create optimal filter weights, $\mathbf{w}_{opt}$, that minimize the least-mean-square of an error signal, $\mathbf{e}(k)$, defined as the difference between $\mathbf{d}(k)$ and the filtered version of $\mathbf{u}(k)$, $\mathbf{y}(k)$. After the calculation of the error for a sample n the weights $\mathbf{w}_{opt}$ are updated and applied to the $k+1$ sample. The equations for the k^{th}

sample are

$$\mathbf{y}(k) = \mathbf{w}_{opt}\mathbf{u}(k) \tag{4.26}$$

$$\mathbf{e}(k) = \mathbf{d}(k) - \mathbf{y}(k) \tag{4.27}$$

$$\mathbf{w}_{opt}(k+1) = \mathbf{w}_{opt}(k) + \mu\mathbf{u}(k)\mathbf{e}(k) \tag{4.28}$$

where μ is a weighting term affecting the convergence rate of the filter to the optimal weights. If μ is very small, the convergence of the filter will be very slow. On the other hand, if μ is too large, the updating of the filter may become unstable. A block diagram of the LMS algorithm is shown in Figure 4.19.

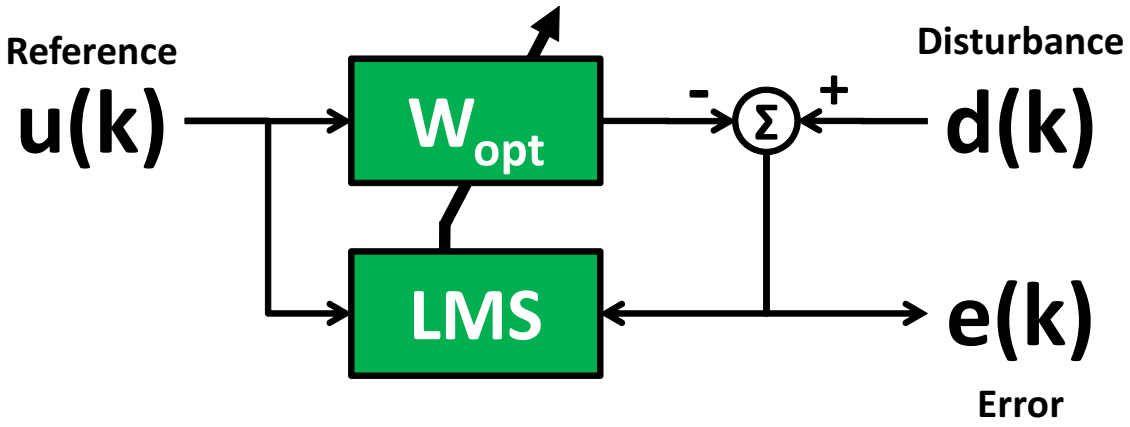
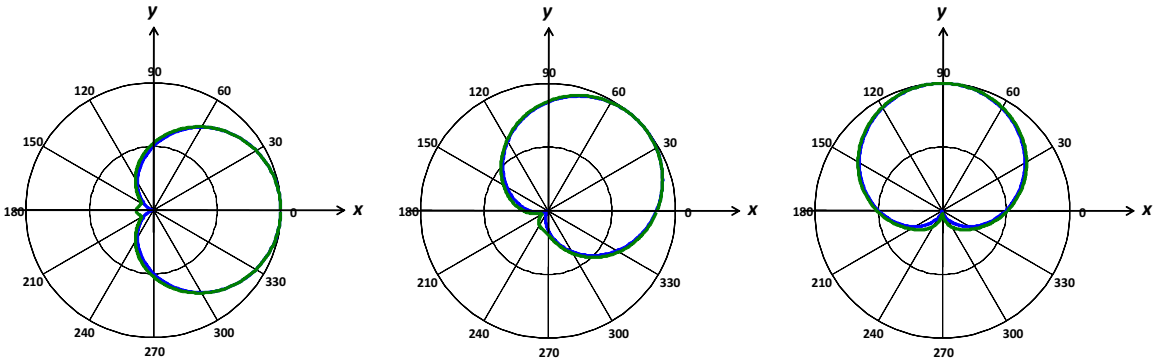


Figure 4.19 Block diagram of LMS algorithm.

4.2 Beampattern Matching

In many cases the measurement of sound with a microphone array may impact the reproduction of sound through speakers or headphones. A study of the impact of different ear coverings on human sound localization is available in [6]. In particular the diffraction of sound around rigid arrays introduces challenges in removing or modifying those diffractive effects to accurately present sounds to listeners. Fortunately there are processing methods available to minimize the difference between a particular set of measured microphone responses and a desired playback environment.

Consider a 3 microphone circular array with equal spacing d between elements, oriented in the $x-y$ plane, and a target cardioid beampattern, oriented at 3 difference azimuth angles, $\theta = 0°$, $45°$, and $90°$, shown in blue in Figures 4.20(a), 4.20(b), and 4.20(c). The beampattern of the filtered circular array response is shown in green in Figures4.20(a), 4.20(b), and 4.20(c). The approximated beampatterns are not perfect matches to the target beampatterns, although in this case, with a fairly simple target beampattern, an increase in the number of equally spaced microphones would improve the match.



(a) $\theta = 0°$ (b) $\theta = 45°$ (c) $\theta = 90°$

Figure 4.20 Target cardiod beampattern (blue) and matched beampattern (green) using a 3 microphone equally spaced circular array oriented in the $x-y$ plane for $kd = \pi/2$.

The relationship between the target array manifold matrix, $\mathbf{D}_{target}$, the array manifold matrix, $\mathbf{D}_{array}$, and the beampattern matching filters, $\mathbf{W}_{BM}$, is

$$\mathbf{W}_{BM}\mathbf{D}_{array} = \mathbf{D}_{target} \tag{4.29}$$

The beampattern matching filters are then calculated by minimizing the squared difference between the target and array manifold matrices by using a matrix inversion,

$$\mathbf{W}_{BM} = \mathbf{D}_{target}\mathbf{D}_{array}^{-1} \tag{4.30}$$

Since the array manifold matrix may represent an unequal number of directions and microphones, the shape of $\mathbf{D}_{array}$ will likely not be square. Therefore the inverse of $\mathbf{D}_{array}$ cannot be carried out as indicated, but rather the pseudo-inverse must be calculated. The

pseudo-inverse is a generalized inverse that can be performed on 2-dimensional matrices of any size, rectangular or square, complex or real,

$$\mathbf{D}_{array}^{+} = (\mathbf{D}_{array}^{H}\mathbf{D}_{array})^{-1}\mathbf{D}_{array}^{H} \tag{4.31}$$

In the event that the matrix $\mathbf{D}_{array}$ is invertible (i.e. square and non-singular), $\mathbf{D}_{array}^{+} = \mathbf{D}_{array}^{-1}$. Since the dimensions of $\mathbf{D}_{array}$ are typically not equal, there actually exist 2 pseudo-inverses, a left-inverse and a right-inverse. In our formulation we are interested in the left-inverse with dimensions [microphones x spatial directions]. Additionally, the problem of $\mathbf{D}_{array}$ being ill-conditioned at low frequencies may prevent the pseudo-inverse from being calculated, requiring diagonal loading (as was seen in section 4.1.3)

$$\mathbf{D}_{array}^{+} = (\mathbf{D}_{array}^{H}\mathbf{D}_{array} + \alpha\mathbf{I})^{-1}\mathbf{D}_{array}^{H} \tag{4.32}$$

As a more complex example, a pictorial version of beampattern matching process is shown in Figure 4.21, where a set of microphone responses are filtered to provide a desired response to the listener. Essentially the sound is filtered to preserve spatial cues so that the listener believes that he/she is really in the original environment. Home theater 5.1, 7.1, 10.1, 11.1, etc. systems are an example of where a single microphone response is used and processed to place sound sources in the correct spatial location for the listener. The opposite of this would be a multi-channel microphone system where signals are processed to present binaural signals to a listener via headphones.

4.2.1 Example: Transparent Hearing

When a person is wearing a set of headphones he/she benefits from passive (and nowadays, possibly active) noise cancellation, but loses nearly all acoustic spatial cues used for localization. The presence of the headphones occludes the pinnae and greatly interferes with the natural HRTF's of the wearer. Synthesis of the natural HRTF's by way of a microphone array mounted on the headphones and a set of signal processing algorithms may sufficiently recreate the auditory cues to the wearer. To be concise, the system will measure and process incident sound and play it back through the headphone

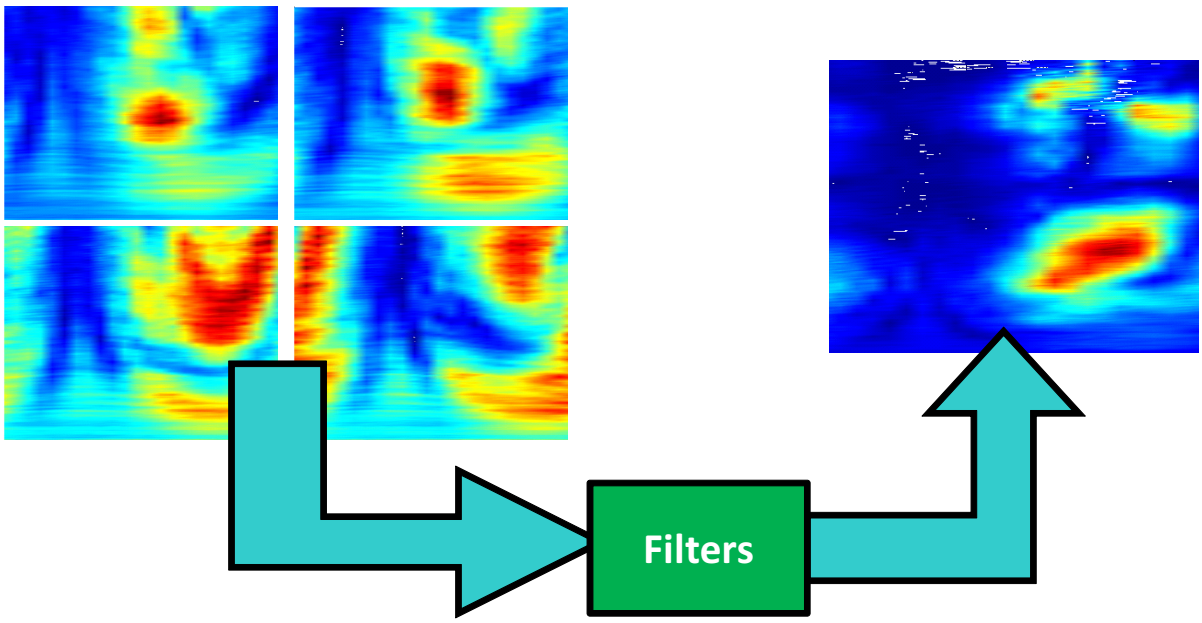


Figure 4.21 Concept of beampattern matching where 4 microphone signals are processed through a set of filters to achieve some desired response to present to a listener.

speakers so that the user feels as if he/she is not wearing headphones at all. This proposed system has been called "transparent hearing," the goal being to make the headphones seem "transparent," or "natural hearing restoration," where the natural hearing of the user is restored by way of the system. Figure 4.22 shows a block diagram of the main concept involved in transparent hearing.

The easiest way to understand the transparent hearing algorithm is by examining the individual physical elements involved in terms of transfer functions. A functional diagram of human hearing without headphones and wearing headphones with the transparent hearing system are shown in Figure 4.23, with the relevant terms described here. The first transfer function of interest is the target HRTF's for the wearer, $\mathbf{D}_{HRTF}$, which can also be thought of as the array manifold matrix for the human ears, shown in the top of Figure 4.23. Without an accurate measurement or estimate of $\mathbf{D}_{HRTF}$ the entire transparent hearing system is difficult to implement. The transfer functions that must be accounted for in the physical system are the array manifold of the headphone array, $\mathbf{D}_{array}$, the occluded HRTF's (what the human ears can still hear with headphones on),

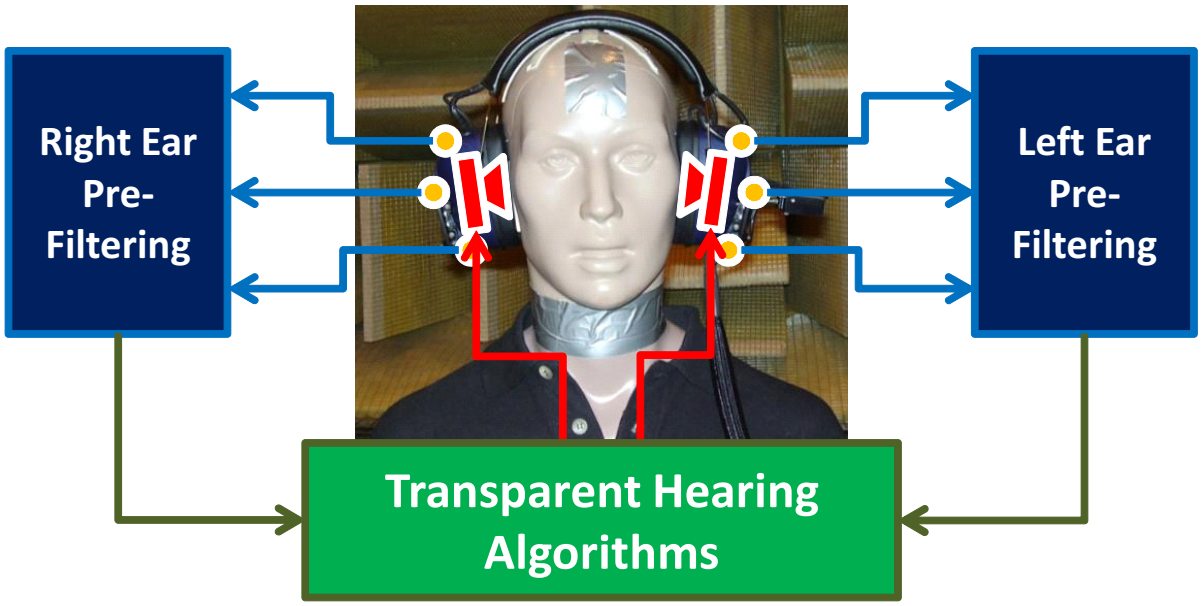


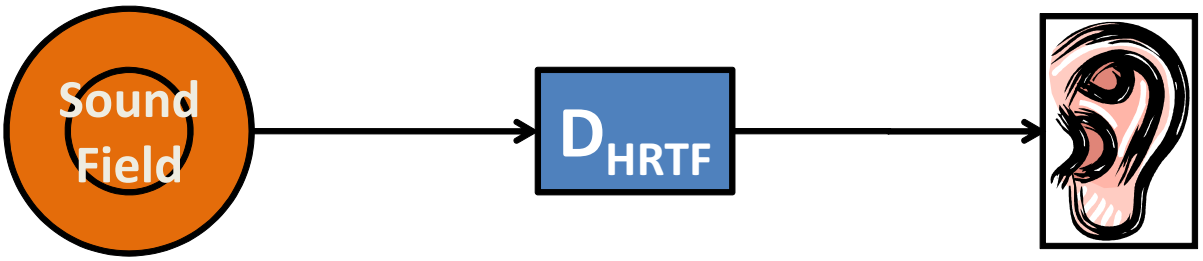
Figure 4.22 Concept of transparent hearing with array microphones (orange), pre-filtering for each headphone (blue), HRTF reconstruction filters and transparent hearing algorithms (green), and signal playback through headphone speakers (red).

$\mathbf{D}_{occ}$, and the transfer functions between the headphone speakers and the ears, $\mathbf{D}_{act}$. At a particular frequency, the dimensions of the transfer function matrices are [spatial directions x microphones] for $\mathbf{D}_{array}$, [spatial directions x ears] for $\mathbf{D}_{HRTF}$ and $\mathbf{D}_{occ}$, [headphones x ears] for $\mathbf{D}_{act}$, and [microphones x headphones] for $\mathbf{W}_{TH}$. The end goal is to create a set of transparent hearing filters, $\mathbf{W}_{TH}$, which will recreate $\mathbf{D}_{HRTF}$ using a combination of microphone responses. For the purposes of calculating $\mathbf{W}_{TH}$, it is assumed that $\mathbf{D}_{array}$ and $\mathbf{D}_{act}$ are constant and that occlusion of the pinnae is perfect, i.e. $\mathbf{D}_{occ}$ is zero, shown in the bottom of Figure 4.23.

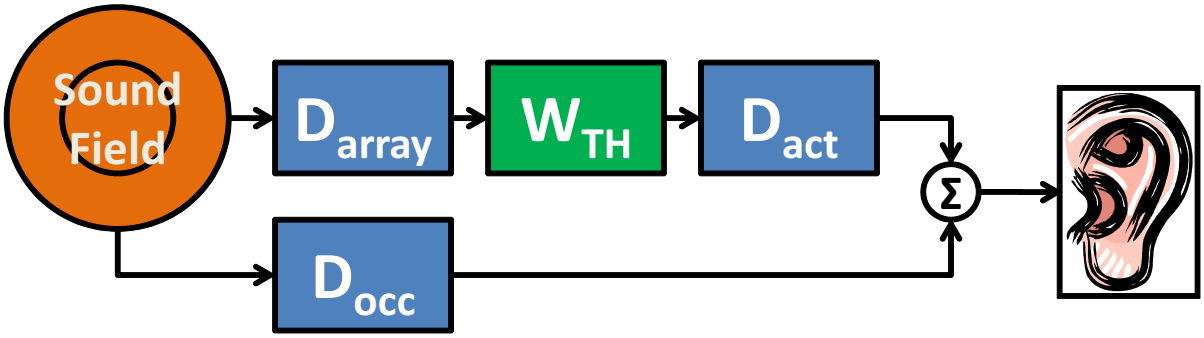
Based on transfer functions described and the diagrams in Figure 4.23, the goal of the transparent hearing filters is, in the frequency domain, to satisfy

$$\mathbf{D}_{array}\mathbf{W}_{TH}\mathbf{D}_{act} = \mathbf{D}_{HRTF} \tag{4.33}$$

If it were just as easy as it is written, the implementation of a transparent hearing system would be trivial. However, there are several problems that must be solved to even make the previous equation realizable. The first problem, inverting a non-square



(a) Natural



(b) Transparent

Figure 4.23 Functional diagrams of natural human hearing and transparent hearing.

(and possibly ill-conditioned) array manifold matrix, was discussed in section 4.2. The next issue is the inversion of $\mathbf{D}_{act}$. Fortunately the dimensions of $\mathbf{D}_{act}$ are [microphones x ears], indicating the relationship between each headphone speaker and each ear. One more simplifying assumption made is that the left headphone speaker has no effect on the right ear and vice versa, i.e. the responses at the left and right sides are uncoupled, causing $\mathbf{D}_{act}$ to be diagonal and invertible. The end result is that the inversion leads to headphone speaker equalizations to account for speaker dynamics in the signal played back to the ears.

In simple terms, $\mathbf{W}_{TH}$ is meant to remove the dynamics associated with the array manifold for the headphone array, transform the sound using the HRTF's for the listener to appropriately recreate magnitude and phase delays associated with spatial auditory cues, and finally apply the reciprocal of the headphone speaker dynamics so that what comes out of the speakers is in the intended response at the ear. A complete calculation of $\mathbf{W}_{TH}$ requires some accommodation for the occluded HRTF's. That is to say that when

the wearer hears a sound, there are actually 2 paths by which the sound gets to the ears. The first path is through $\mathbf{D}_{occ}$, where the sound has been attenuated and the delays have been changed. The second path is through the transparent hearing system, where ideally all of the correct cues have been restored. Presenting the recreated sound at or before the time sound arrives through the headphones is physically not realizable. Addressing this very real situation is the following: if the measurement, processing, and presentation of the recreated sound is within the 70ms window dictated by the precedence effect, the system can avoid the appearance of an echo. The optimal solution to Equation 4.33, assuming no terms from Figure 4.23(b) can be neglected and that all necessary inversions are possible, is

$$\mathbf{W}_{TH} = (\mathbf{D}_{array}^{H}\mathbf{D}_{array} + \alpha\mathbf{I})^{-1}\mathbf{D}_{array}^{H}[\mathbf{D}_{HRTF} - \mathbf{D}_{occ}]\mathbf{D}_{act}^{-1} \tag{4.34}$$

Under the assumption that the occluded HRTF's can either be neglected or their effect can be treated with an independent system, the formulation for the optimal transparent hearing filter is simplified to the form

$$\mathbf{W}_{TH} = (\mathbf{D}_{array}^{H}\mathbf{D}_{array} + \alpha\mathbf{I})^{-1}\mathbf{D}_{array}^{H}\mathbf{D}_{HRTF}\mathbf{D}_{act}^{-1} \tag{4.35}$$

Neglecting the occluded HRTF's has much to do with the difficulty of measuring $\mathbf{D}_{occ}$. Headphone position, orientation, cup sealing (or lack thereof) all affect the true occluded HRTF's that a wearer experiences. The dynamic nature of $\mathbf{D}_{occ}$ may dictate an adaptive approach to deal with its effects.

With all of the theory discussed, the reality is that the implementation of a suitable transparent hearing system will likely be tailored to each wearer. The perceptual abilities of human hearing vary from person to person and the performance evaluation of a system affecting human perception is, by its very nature, subjective. Humans develop perception based on lifetime experiences, so at the very least if the wearer is given time to "learn the system," he/she may adequately adjust to whatever initial foibles they perceive.

CHAPTER 5

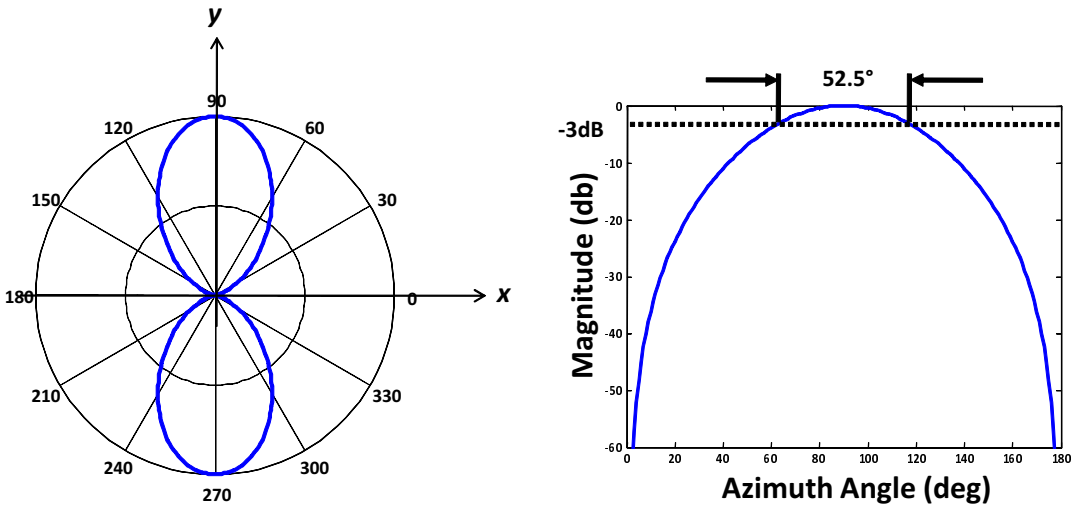
Array Performance

While there are many array geometries and processing methods to choose from for any application, the deciding factor in what geometry and method to choose is based on performance. This chapter presents some of the numerous ways to quantify the performance of microphone arrays and their associated filtering schemes. The merits of different array geometries will be discussed by keeping beamforming methods constant, i.e. comparing line, circular, and spherical arrays. Descriptions of beamformer performance will be presented by comparing the relative sizes of the main and side lobes. The performance gain from mounting microphone arrays on diffracting shapes will be shown using metrics for directional gain, sensitivity to uncorrelated noise, and singular values. A helmet mounted array will be shown to have improved performance over a spherical array of comparable size. Finally, the real-world localization performance of the helmet array will be demonstrated through experiments involving helicopters, automobiles, and gunshots.

5.1 Beampattern

A beampattern is a way to simultaneously view the gain in the desired look direction and the suppression of signals from all other (assumed undesired) directions. Within a beampattern there are several parameters that indicate the performance of a particular array with a particular beamformer applied. Figure 5.1 shows 2 representations of the

same beampattern for a 20 element equally spaced line array in broadside configuration for a kd value of $\pi/10$. Figure 5.1(a) shows a polar beampattern with linear magnitude while Figure 5.1(b) shows beampattern magnitude in dB for angles between 0° and 180° (the range between 180° and 360° is omitted because of redundancy due to symmetry). Beamwidth, normally called -3dB beamwidth, is defined as the width of the main lobe between the points at which the directional gain drops 3dB from its maximum, as indicated in Figure 5.1(b). In this example the beamwidth is 52.5° in the azimuth plane. The suppression of side lobes is measured between the maximum gain of the main lobe and the maximum gain of any side lobes. The same line array is used with a minimum variance beamformer to yield the beampatterns shown in Figure 5.2. Finally, both beampatterns are shown in Figure 5.3 for a direct comparison. The beampatterns shown in

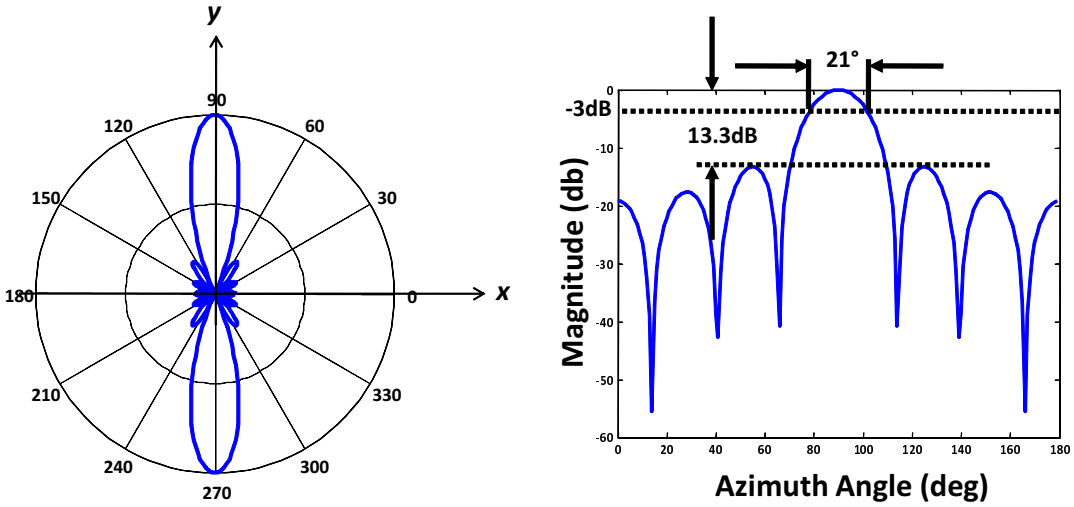


(a) Linear scale

(b) dB scale

Figure 5.1 Beampattern for 20 microphone line array with d spacing for a kd value of $\pi/10$ in broadside configuration using conventional beamformer.

Figure 5.3 provide an illustration of the differences between 2 beamforming strategies. The minimum variance beamformer clearly has many more nulls than the conventional beamformer, and on the average suppression from other directions is better as well. Not apparent in Figure 5.3 is the susceptibility of the array to uncorrelated noise at the microphones. In the case of the minimum variance beamformer, a compromise has been made



(a) Linear scale (b) dB scale

Figure 5.2 Beampattern for 20 microphone line array with d spacing for a kd value of $\pi/10$ in broadside configuration using a minimum variance beamformer.

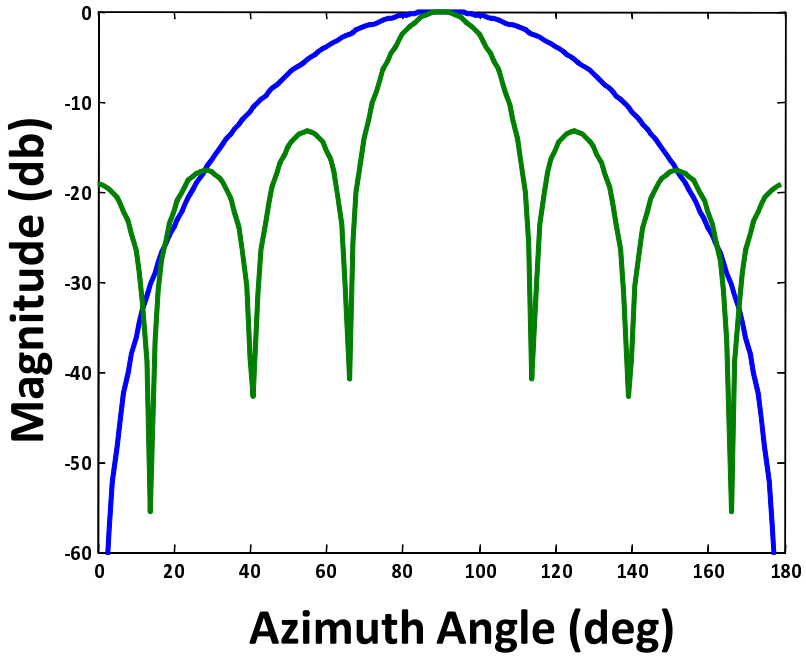


Figure 5.3 Direct beampattern comparison of conventional beamformer (blue) and MVDR beamformer (green).

with this susceptibility to achieve greater directivity. In the case of the conventional beamformer, the susceptibility is kept at a minimum, as will be discussed later.

5.2 Directivity

A commonly used metric for describing the performance of a microphone array is the directivity index (DI), or the ability of an array to suppress a diffuse noise field. The directivity index is calculated by dividing the power in the look direction by the power in all directions using the diffuse noise field CSD [69], $\mathbf{\Phi_{xx}}$. The directivity index integral can be numerically approximated by,

$$DI = 10\log_{10}\frac{|\mathbf{w}_\theta^H\mathbf{d}_\theta|^2}{\mathbf{w}_\theta^H\mathbf{\Phi_{xx}}\mathbf{w}_\theta} \tag{5.1}$$

where the calculation of a noise field CSD in general and for the diffuse case is discussed in section 4.1.3.1.

An easier way to view directivity is with beampatterns. From section 5.1, Figure 5.1(a) showed a beampattern with low directivity while Figure 5.2(a) showed a beampattern with high directivity. Since the purpose of the MVDR beamformer is to maximize the directivity of the beampattern, it should come as no surprise that the MVDR beampattern (Figure 5.2(a)) had a higher directivity than the conventional beamformer (Figure 5.1(a)).

Calculation of the minimum variance beamformer filter weights may require diagonal loading of the CSD matrix to condition it for inversion. The amount of diagonal loading added to the matrix directly impacts the directivity, and as we will later see, the sensitivity of the beamformer to uncorrelated noise at the microphones. Restating Equation 4.19

$$\mathbf{w}_\theta = \frac{[\mathbf{\Phi_{xx}} + \alpha\mathbf{I}]^{-1}\mathbf{d}_\theta}{\mathbf{d}_\theta^H[\mathbf{\Phi_{xx}} + \alpha\mathbf{I}]^{-1}\mathbf{d}_\theta} \tag{5.2}$$

where α is the coefficient changed to add more or less diagonal loading. The value of α is highly dependent upon the norm of $\mathbf{\Phi_{xx}}$, necessitating a normalized form of Equation 5.2. Rather than using the CSD matrix, the coherence matrix $\mathbf{\Gamma_{vv}}$ has unity entries

on the diagonal, allowing an assessment of diagonal loading values between different beamformers. The calculation of the coherence matrix is given by

$$\mathbf{\Gamma}_{\mathbf{v}_n\mathbf{v}_m} = \frac{\mathbf{\Phi}_{\mathbf{x}_n\mathbf{x}_m}}{\sqrt{\mathbf{\Phi}_{\mathbf{x}_n\mathbf{x}_n}\mathbf{\Phi}_{\mathbf{x}_m\mathbf{x}_m}}} \tag{5.3}$$

and so the calculation of the optimal filter with the coherence matrix is

$$\mathbf{w}_\theta = \frac{[\mathbf{\Gamma}_{\mathbf{vv}} + \epsilon\mathbf{I}]^{-1}\mathbf{d}_\theta}{\mathbf{d}_\theta^H[\mathbf{\Gamma}_{\mathbf{vv}} + \epsilon\mathbf{I}]^{-1}\mathbf{d}_\theta} \tag{5.4}$$

The effect of this loading on the beampattern is shown in Figure 5.4. The number of microphones used for Figure 5.4, 20, can be seen as the number of nulls present in the beampattern corresponding to $\epsilon = 10^{-10}$. Notice that as ϵ is increased the main lobe beamwidth increases, the number of side lobes decreases, and in the limit of large ϵ, a conventional beamformer beampattern is reached. Presenting the relationship between ϵ and directivity in another way, Figure 5.5 shows directivity index as a function of ϵ, where the labels "SDB" and "DSB" show the maximum and minimum values of the directivity index achieved by using a superdirective beamformer or a delay-and-sum beamformer, respectively.

5.3 Robustness

Another metric for describing array performance is the ability of the array to suppress uncorrelated noise between the sensors, generally caused by noise within the sensors. This suppression of uncorrelated noise is called the white noise gain (WNG) of the array beamformer. The calculation of the WNG is the same as for the DI of the array, but with the diffuse noise field CSD matrix, $\mathbf{\Phi}_{\mathbf{xx}}$, replaced by the CSD matrix for uncorrelated noise, the identity matrix $\mathbf{I}$,

$$WNG = 10\log_{10}\frac{|\mathbf{w}_\theta^H\mathbf{d}_\theta|^2}{\mathbf{w}_\theta^H\mathbf{I}\mathbf{w}_\theta}) = 10log_{10}(\frac{|\mathbf{w}_\theta^H\mathbf{d}_\theta|^2}{\mathbf{w}_\theta^H\mathbf{w}_\theta} \tag{5.5}$$

Equation 5.5 is suitable for the form of $\mathbf{d}$ that results from the free field case. In the non-free field case (i.e. diffracting), the form of Equation 5.5 changes to,

$$WNG = 10\log_{10}\frac{|\mathbf{w}_\theta^H\mathbf{d}_\theta|^2}{\mathbf{d}_\theta^H\mathbf{d}_\theta\mathbf{w}_\theta^H\mathbf{w}_\theta} \tag{5.6}$$

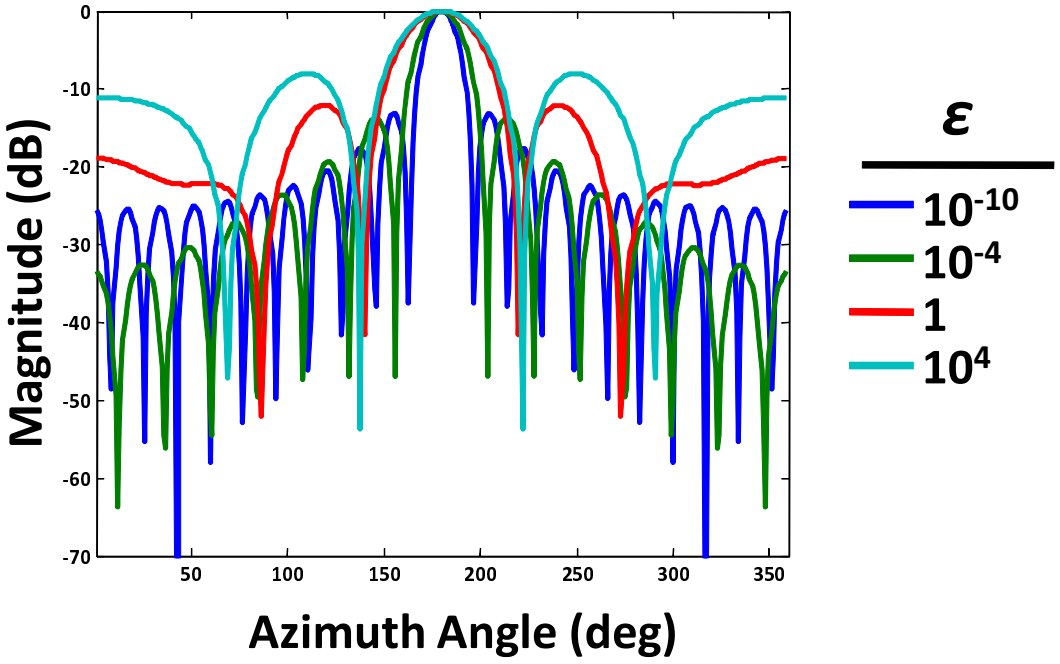


Figure 5.4 Effect of diagonal loading on beampattern directivity for a 20 microphone circular array with equal spacing d between elements at a kd value of $\pi/3$.

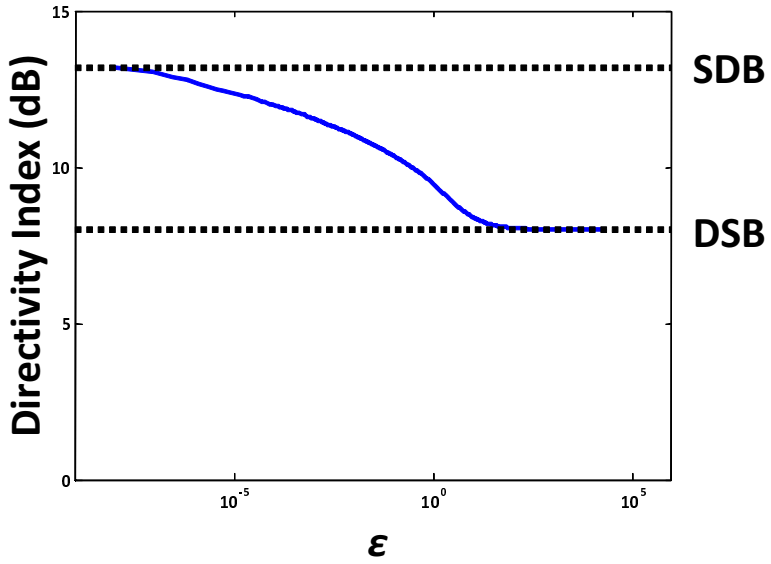


Figure 5.5 The effect of diagonal loading on the directivity index of an array beamformer ($kd = \pi/3$).

where the additional term in the denominator, $\mathbf{d}_\theta^H \mathbf{d}_\theta$, is necessary to account for diffraction effects (i.e. non-unit amplitudes measured at the sensors).

The maximum value of the WNG is governed by the number of sensors used in the array, and the maximum occurs when a conventional beamformer is used. Using the same 20 microphone equally spaced circular array example from the previous section, the relationship between white noise gain and ϵ is shown in Figure 5.6. Recalling Figure

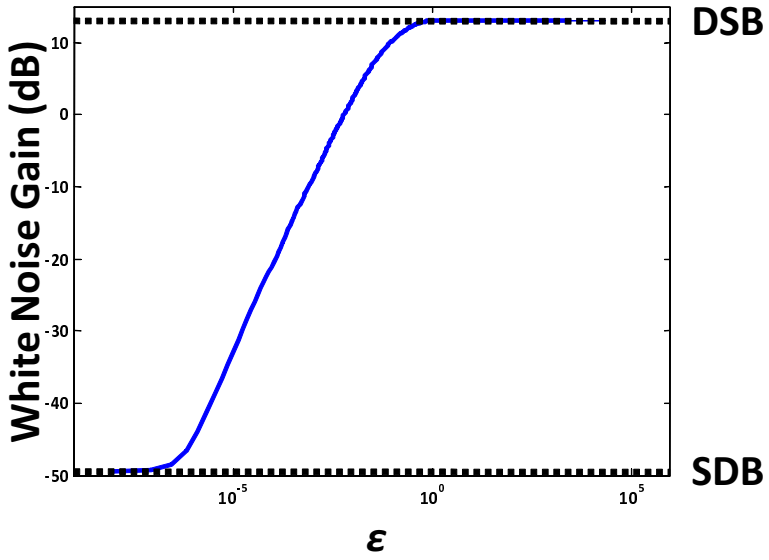


Figure 5.6 The effect of diagonal loading on the white noise gain of an array beamformer ($kd = \pi/3$).

5.5 from the previous section, there is a tradeoff present between the directivity index and the white noise gain of the beamformer, based on the amount of diagonal loading introduced by ϵ. The compromise between directivity and white noise gain is shown in Figure 5.7. For this example, notice the steep decline in white noise gain for relatively small increases in directivity, indicating very high susceptibility to sensor noise in exchange for better suppression of spatially diffuse noise. In applications where high quality, low noise microphones are implemented, a fairly low white noise gain may be acceptable to achieve a desired directivity. As the operating frequency of the beamformer is increased, the tradeoff between the directivity and white noise gain of the beamformer becomes less and less important, a function of the CSD matrix becoming better and better con-

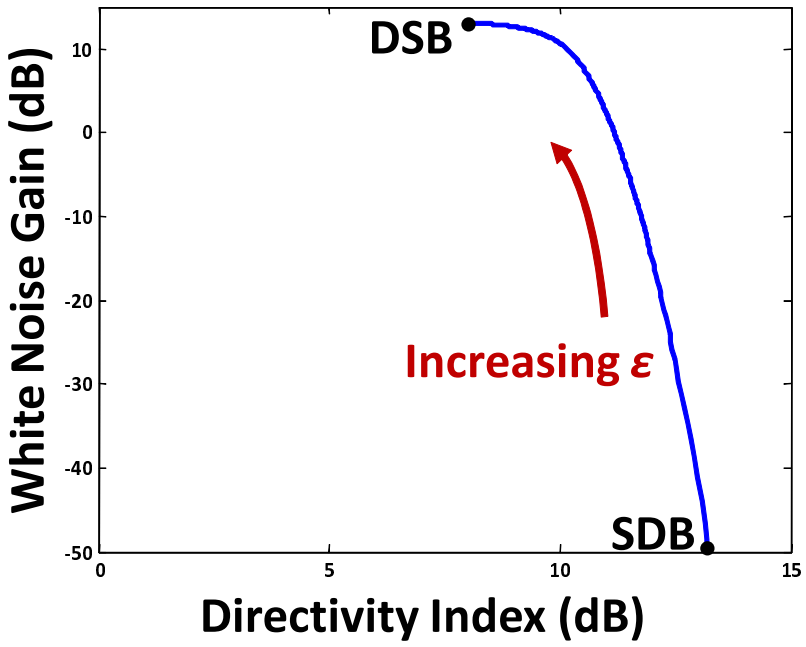


Figure 5.7 Tradeoff between white noise gain and directivity of a beamformer ($kd = \pi/3$).

ditioned. Figure 5.8 shows the white noise gain/directivity tradeoff curves for several wavenumbers for the same range of ϵ. Notice that as the wavelength of the incident sound becomes comparable to the spacing of the microphones ($kd = 2\pi$) the impact of diagonal loading becomes insignificant. The reason for the decrease in directivity as the wavelength decreases further is due to introduction of phase uncertainty as the microphone spacing is approaching multiple sound wavelengths. The shapes of the white noise gain/directivity curves has significance with respect to spherical harmonics, something that will be discussed in the next section.

5.4 Singular Value Decomposition

The singular value decomposition (SVD) can be used to examine the CSD matrix and the effect of diagonal loading on directivity and robustness to uncorrelated noise. SVD factors an input matrix into the product of two unitary matrices and a diagonal

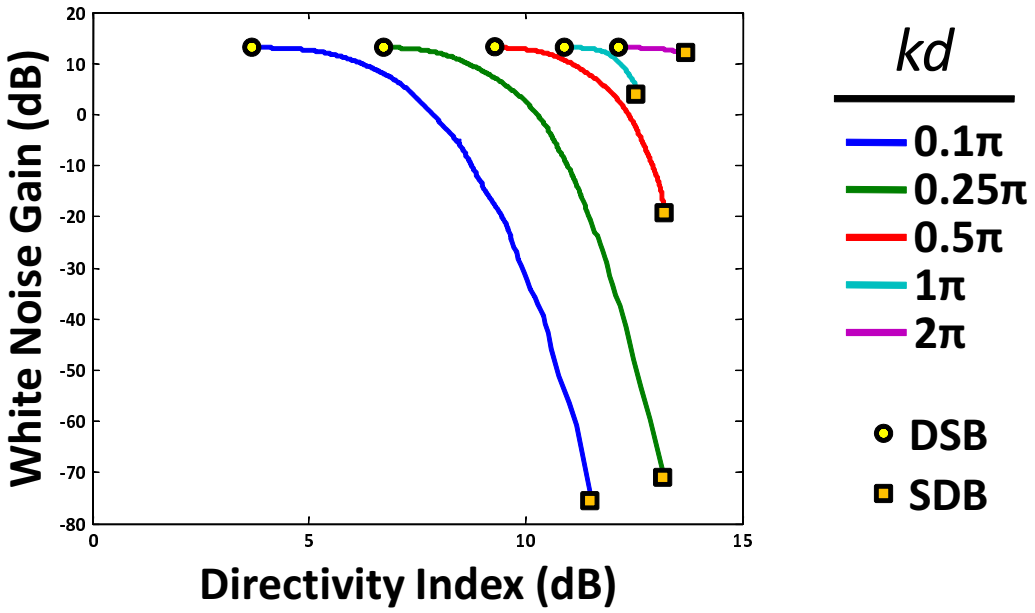


Figure 5.8 Impact of decreasing wavelength on white noise gain/directivity index trade-off.

matrix [73]. Using the same CSD matrix, $\mathbf{\Phi_{xx}}$, that was used previously,

$$\mathbf{\Phi_{xx}} = \mathbf{U\Sigma V}^H \tag{5.7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary, and $\mathbf{\Sigma}$ has the form

$$\mathbf{\Sigma} = diag(\sigma_1, \sigma_2, \ldots, \sigma_n) \tag{5.8}$$

where n is equal to the minimum dimension of the array manifold matrix, $\mathbf{D}$, and σ_1, σ_2, $\ldots$, σ_n are the singular values of $\mathbf{\Phi_{xx}}$. Examining the singular values of $\mathbf{\Phi_{xx}}$ shows how each contributes to array performance at a particular ka value. The contribution of each singular value can be determined by examining its magnitude relative to other singular values. At low values of ka there will be one relatively large singular value, while at higher ka values the magnitude of all singular values will be comparable. The dominant singular value at low ka values is related to the negligible phase differences between array microphones, meaning that compared to a wavelength, the array aperture is essentially a point. Diagonal loading of $\mathbf{\Phi_{xx}}$ at these low ka values adds noise to $\mathbf{\Phi_{xx}}$, drowning the contributions from smaller singular values in noise. This artificial noise has the effect of

decreasing the DI and increasing the WNG of the array beamformer. As stated in section 5.3, increased WNG can be achieved by increasing the number of microphones within the array, regardless of the array geometry. In contrast, increased DI depends upon the array geometry, number of microphones, noise field properties, and ka value.

Consider a spherical microphone array and a set of spherically isotropic sources in the far-field. In the limit of an infinite number of evenly spaced microphones and sources, the left and right singular vectors represent the spherical harmonics (introduced in section 2.3.2). These singular vectors can be thought of as generalized radiation modes [74]. For a finite number of microphones, the singular vectors are essentially sampled versions of the spherical harmonics. Equivalently, the singular values can be thought of as the radiation efficiencies of each of the spherical harmonics at a particular value of ka. That is to say that the relationship between the sources and receivers (microphones) is equivalent to replacing all sources with receivers and vice versa. These phenomenon have been studied in detail in the reciprocal relationship of radiation efficiency modes in [75]. For a non-spherical array the radiation modes are not exactly the spherical harmonics and can change shape with frequency.

It is important to note that the spherical harmonics themselves are independent of ka, but that the radiation efficiencies of each harmonic are frequency dependent, as shown by the normalized singular values in Figure 5.9. The representations of the spherical harmonics from Figure 2.15 have been placed over their corresponding singular values (i.e. 1 monopole, 3 dipoles, and 5 quadrupoles). Figure 5.9 shows an approximation of the normalized singular values using a 266 microphone spherical array and 366 far-field sources.

5.5 Array Comparisons

With the underlying theory behind microphone arrays and performance metrics sufficiently introduced, some comparisons between different array types can be made. Analytical, numerical, and experimental results will be compared for various geometries to

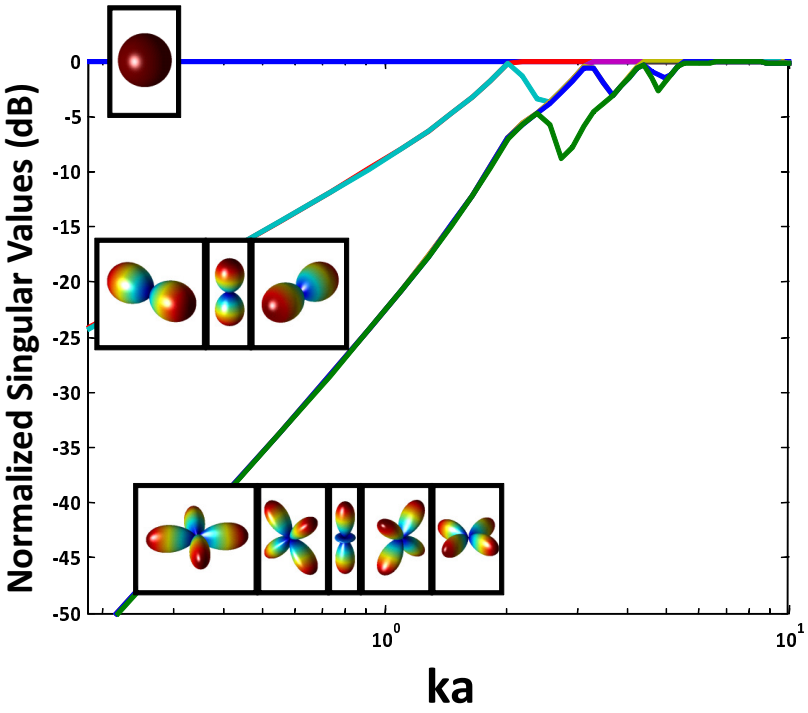


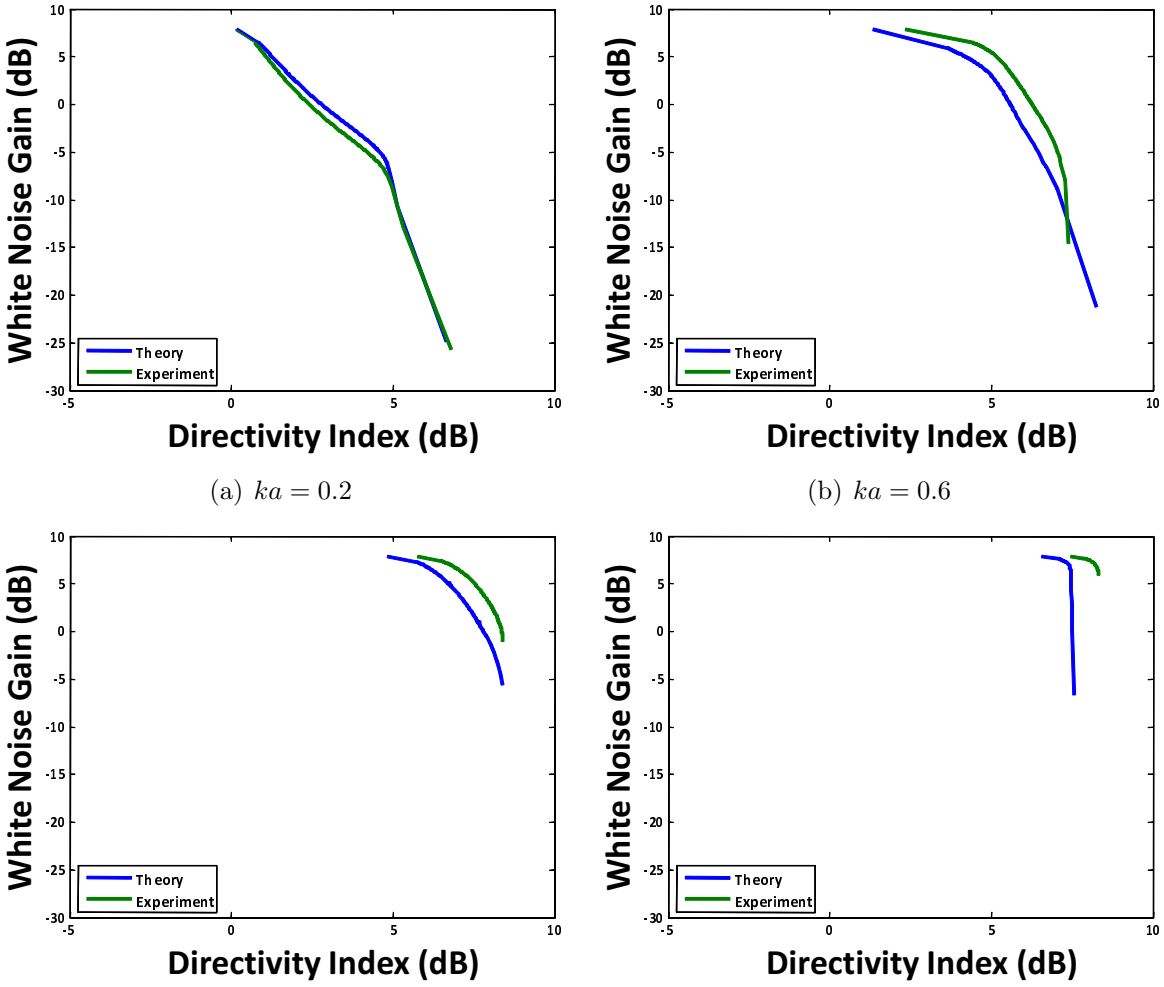
Figure 5.9 Normalized singular values and associated spherical harmonics for a free field spherical array.

show consistency. Free field and diffracting geometries will also be compared to show the benefit of using diffracting mounting bodies for microphone arrays.

5.5.1 Cylinder

For the purposes of comparison, a 6 element equally spaced circular microphone array with a diameter of 0.2m, approximately the diameter of the human head, will be used in this section. Also, the set of directions used for the array manifold is limited to 24 equally spaced angles in the azimuth plane. The presence or absence of a diffracting mounting body will have an impact on the performance of the array, and that impact can be viewed in terms of the directivity index, white noise gain, and ultimately, singular values. The first comparison made is a verification of experimental measurements against theoretical values for a free field array. Figure 5.10 shows the white noise gain and directivity indexes for the circular array at four values of ka: (a) 0.2, (b) 0.6, (c) 1.2, and (d) 2.4. For Figures

5.10, 5.11, and 5.12, the "look" direction is at an azimuth angle of 0° and an elevation angle of 0°.

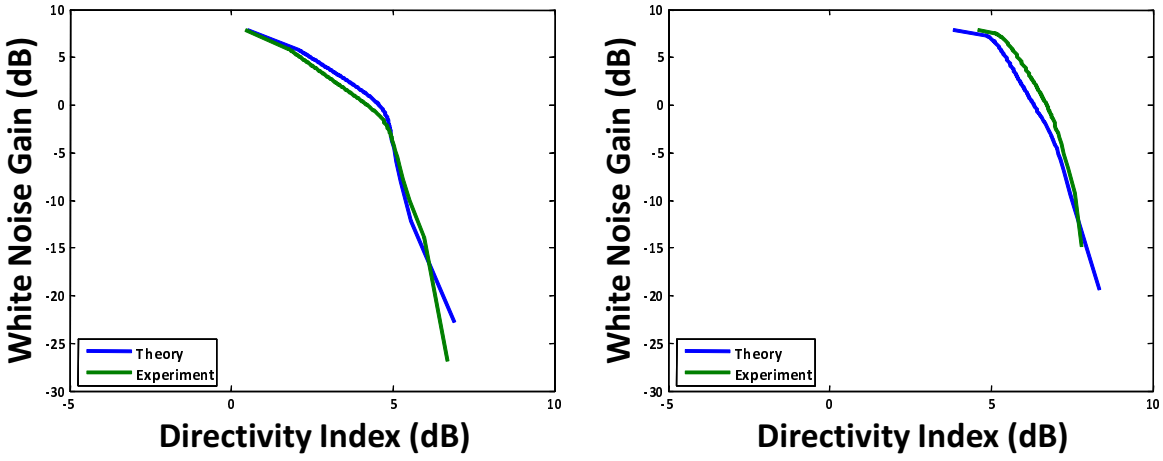


(a) $ka = 0.2$ (b) $ka = 0.6$

(c) $ka = 1.2$ (d) $ka = 2.4$

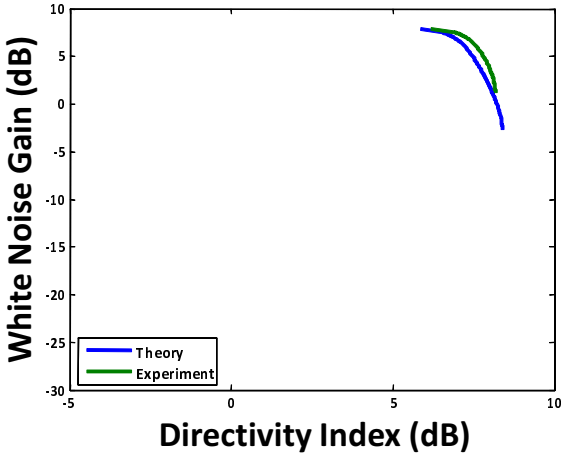
Figure 5.10 Directivity indexes and white noise gains for an equally spaced 6 microphone circular array in the free field from theory (blue) and experimental measurement (green) for several values of ka in the 0° azimuth, 0° elevation direction.

Using Figure 5.10 as a template, Figure 5.11 shows the directivity indexes and white noise gains for a diffracting array of identical geometry to the one previously introduced, for theory and experiment. Notice the same close agreement in the diffracting case as was seen in the free field case, indicating expectation from theory is seen in real world measurements.

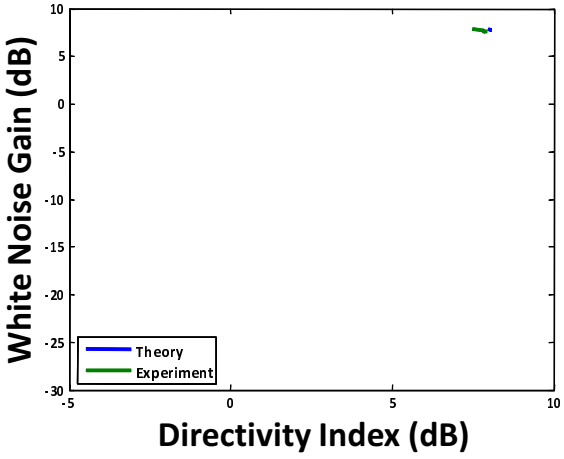


(a) $ka = 0.2$

(b) $ka = 0.6$



(c) $ka = 1.2$



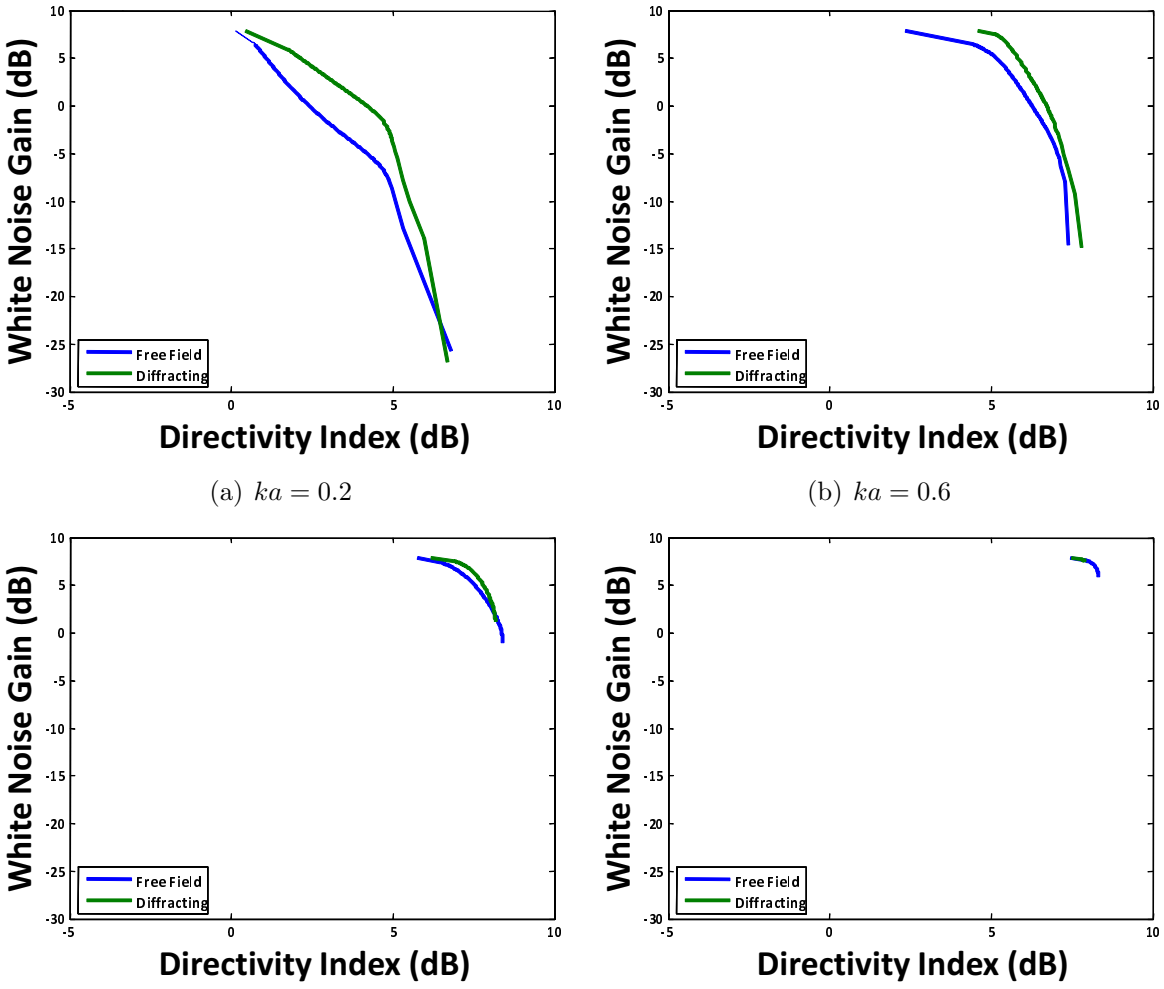
(d) $ka = 2.4$

Figure 5.11 Directivity indexes and white noise gains for an equally spaced 6 microphone circular array mounted on a diffracting cylinder from theory (blue) and experimental measurement (green) for several values of ka in the 0° azimuth, 0° elevation direction.

With good agreement between theory and experiment the comparison between free field and diffracting arrays of identical geometries can be made. Figure 5.12 shows the directivity indexes and white noise gains for the free field and diffracting array experimental measurements for the same values of ka used previously. Since high directivity and high white noise gain are the desirable features, the optimal performance is in the upper right corner of each plot in Figure 5.12. The dominant feature in Figure 5.12 is that the diffracting array always outperforms the free field array. For example, at $ka = 0.2$ the white noise gain is improved by 6dB at a directivity index of 5dB. The improved performance comes from diffraction effects causing an increased travel time for acoustic waves, thereby increasing the phase differences between microphones and essentially increasing the array aperture [76].

The presence of "knees" or "bends" at low ka values in Figures 5.10, 5.11, and 5.12 is due to the contributions from different sets of singular values. When matrix conditioning is at a maximum (maximum white noise gain) only one singular value (corresponding to a monopole) contributes to the beamformer performance. As matrix conditioning is decreased (decreasing white noise gain and increasing directivity index) an additional pair of singular values (corresponding to 2 dipoles) improve beamformer performance. For this example of a 6 element array the remaining 3 singular values (corresponding the last dipole and 2 of 5 possible quadrupoles) continue to improve performance as conditioning is moved to the superdirective beamformer case. The tradeoff between directivity index and white noise gain exists because smaller singular values (i.e. from quadrupoles) increase performance less than larger singular values (i.e. from dipoles). This is seen for lower ka values in Figures 5.10, 5.11, and 5.12 as the different slopes in the curves before and after the "knees."

While Figure 5.12 shows the performance benefit of diffraction at a few values of ka, the singular values illustrate the benefit over the entire frequency range. The larger the singular values are at a particular frequency, the better the performance potential of an array and beamformer. Figure 5.13 shows the normalized singular values for the free field and diffracting arrays presented previously. The important feature in Figure 5.13



(a) $ka = 0.2$

(b) $ka = 0.6$

(c) $ka = 1.2$

(d) $ka = 2.4$

Figure 5.12 Directivity indexes and white noise gains for an equally spaced 6 microphone circular array from experimental measurement for free field (blue) and diffracting (green) arrays for several values of ka in the 0° azimuth, 0° elevation direction.

is that at any frequency the singular values for the diffracting array case are larger than the singular values for the free field case.

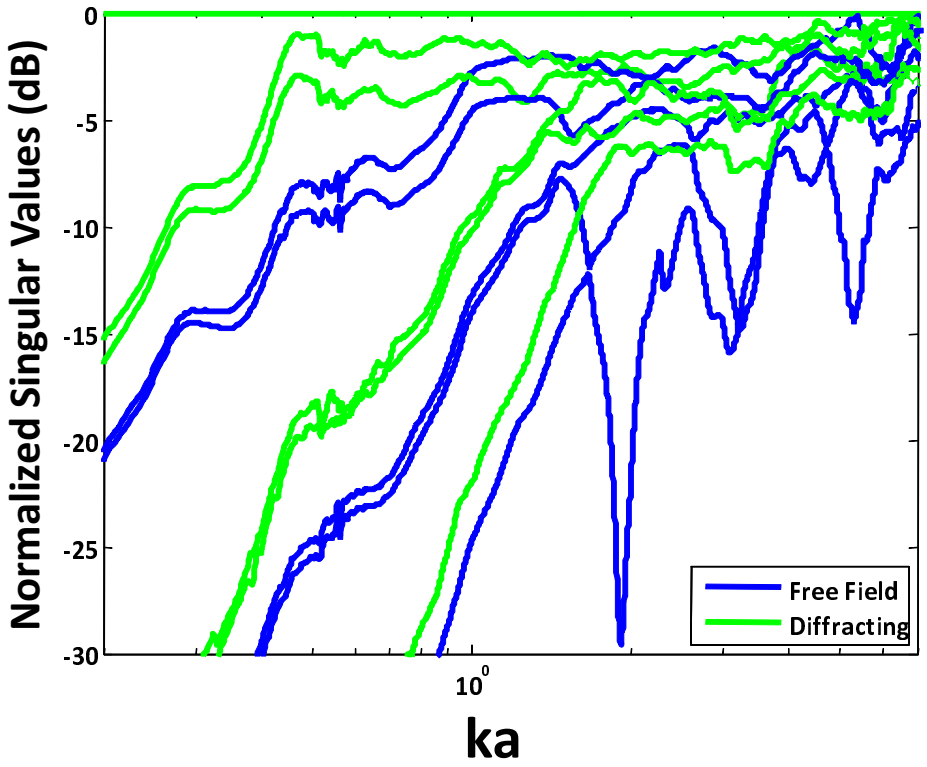


Figure 5.13 Normalized singular values for a circular array in the free field (blue) and on a diffracting cylinder (green).

At higher ka values both of the cases are well conditioned and the singular values are all quite large, but even so the diffracting case is superior to the free field case. Notice that at $ka = 0.2$ the dipole singular values for the diffracting array case are 5dB higher than for the free field case. At $ka = 0.2$, the diameter of the free field array would need to be 55% larger than the diffracting array to achieve the same performance, illustrated in Figure 5.14 by comparing the 0.2m diameter diffracting array to a 0.31m diameter free field array.

What the previous examples have shown is that the presence of a diffracting cylinder improves every performance measure presented thus far. The mechanism by which performance is improved is increased phase delays between microphones caused by acous-

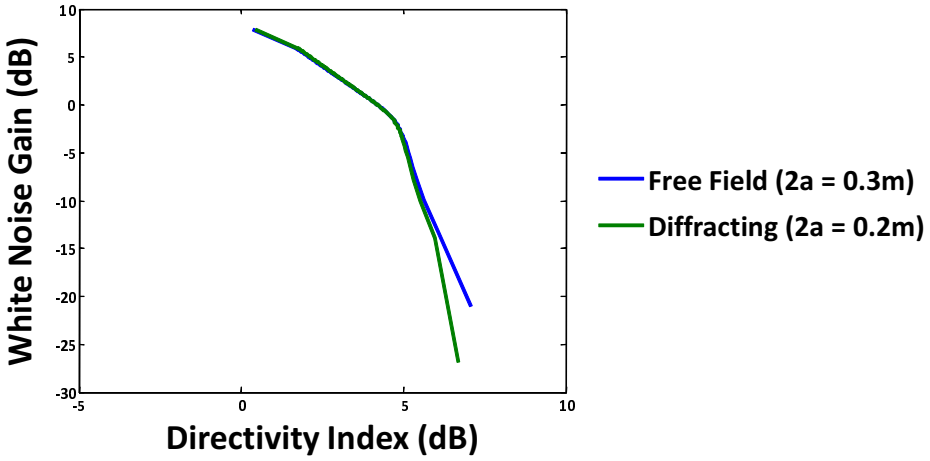


Figure 5.14 Comparison of performance at $ka = 0.2$ for a free field array of radius 0.155m (blue) and a diffracting array of radius 0.1m (green).

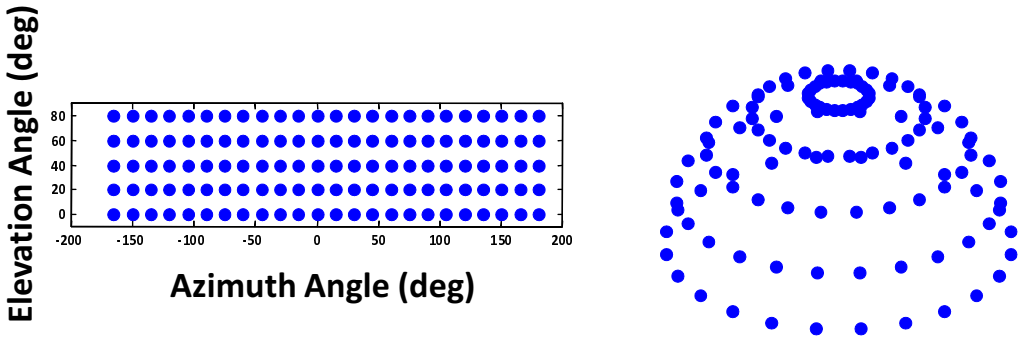
tic diffraction around the cylinder. Regardless of the beamforming strategy, diffraction affords greater potential performance.

5.5.2 Sphere

Extending the comparison of free field arrays to diffracting arrays into the realm of 3-dimensional arrays, a spherical array design will be analyzed. A 6 microphone arrangement will be considered, with 3 microphones placed on the equator of the sphere and 3 microphones placed at an elevation angle of 55° above the azimuth plane. The spacing strategy used is meant to have low sensitivity to sounds in the $-z$ direction (the assumption that sounds do not generally originate at a person's feet). This strategy is chosen to imitate a helmet array application where the shape of the helmet dictates that microphones be placed at or above the equator level. Examples of the diffracting arrays used for experimental measurement were shown in Figure 3.6. The source directions used for calculating the array response were in 15° increments in the azimuth plane and 20° increments in the elevation plane from 0° to 80°, for a total of 120 source directions. The result of equally spacing the sources in both the azimuth and elevation directions is an unequal allocation of area for each source direction. This phenomenon can best

be compared to geography, where a curved surface (i.e. the earth) is projected onto a 2-dimensional rectangular representation, known as a Mercator projection.

In the case of source directions, the sources are essentially placed based on a rectangular grid and then wrapped around a sphere. The problem caused is that points at the equator are spread apart much more than points at the poles, leading to a heavy weighting of responses at the poles and a much lower weighting of the points on the equator. Figure 5.15 shows the phenomenon with a rectangular grid of points and those points wrapped around a sphere. Notice how the sources at the pole of the sphere are much more closely spaced together compared to the spacing at the equator.



(a) Rectangular grid (b) Spherical projection

Figure 5.15 Comparison of rectangular and spherical mapping of grid points.

The closer the sources are together, the more similar their contributions will be to the array response. Therefore, if the spacing strategy in Figure 5.15 were employed without any compensation for the unequal area distribution of the points, the calculation of beamformers would be heavily biased to the $+z$ direction. The matrix $\mathbf{B}$ must be used to account for the relatively coarse sampling near the equator and very dense sampling near the poles. Therefore b_{mm}, the m^{th} diagonal element of $\mathbf{B}$ corresponding to the m^{th} angle, is weighted in proportion to the square root of the cosine of its elevation angle.

$$\mathbf{b}_{mm} = \sqrt{\cos \phi_m} \tag{5.9}$$

Applying the weighting matrix to the array manifold, the form of the CSD for the array changes to

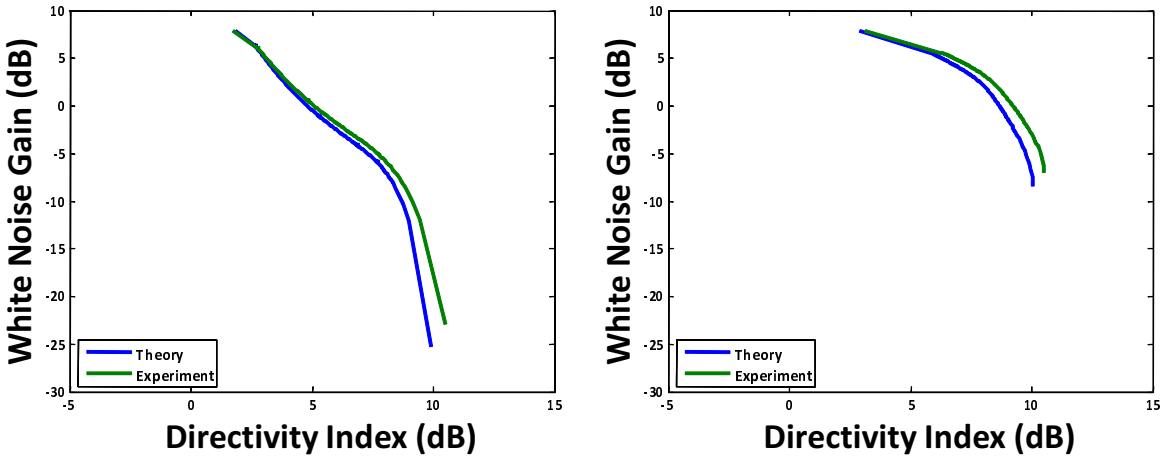
$$\mathbf{\Phi_{xx}} = \mathbf{BD^H DB} \tag{5.10}$$

Alternatively, a source direction strategy could be used that equally spaces all sources, but as in the case of microphone positions within a spherical array, the geometries would again be limited to the Platonic solids. Any other locating strategy will inherently require a weighting matrix to account for the unequal distribution of sources. Using the same progression from the previous section, the first comparison made is the directivity indexes and white noise gains for the theoretical and experimentally measured free field spherical array, shown in Figure 5.16. For the most part the agreement is very close, with any disparities most likely due to the self noise of the measurement equipment and system for the experimental measurements.

The comparison of theoretical and experimental diffracting results also shows a good match, shown in Figure 5.17, with a slight allowance made for the disparity between ideal and real-world measurements. As in the case of the circular array mounted on a cylinder, once the wavelength of incident sound is comparable to the spacing of the microphones, the directivity and white noise gain are approximately equal for any beamformer between conventional and minimum variance.

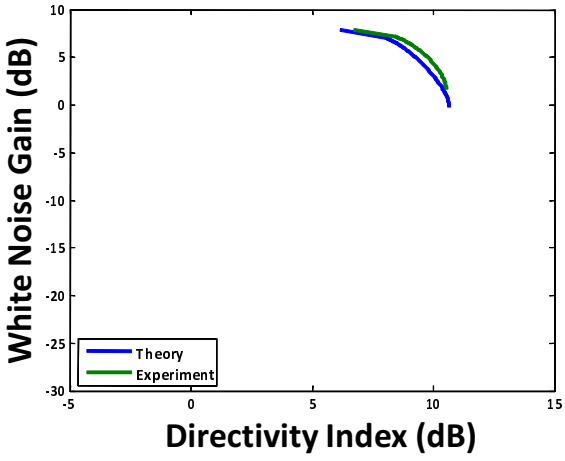
Having established a good agreement between theory and experiment, the final performance comparison to make is between the free field and diffracting spherical arrays. The directivity index and white noise gain values for these arrays are shown in Figure 5.18. While not as drastic as the circular arrays, the performance benefit of a diffracting sphere within the spherical array is still evident. The low frequency performance is where the greatest gains are visible, and this is a common problem area for microphone arrays. For example, at $ka = 0.2$ the white noise gain is 3dB higher at a directivity index of 5dB for the diffracting case over the free field case.

The relationship between the singular values and the tradeoff between directivity index and white noise gain, discussed in section 5.5.1, is shown in Figures 5.16, 5.17, and 5.18. The "knees" at low ka values in these Figures again represent points where the

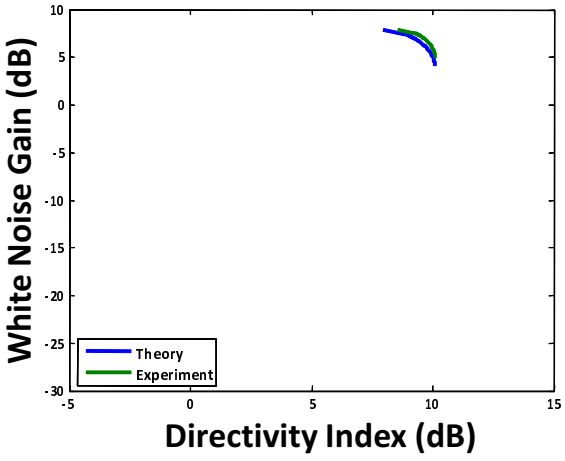


(a) $ka = 0.2$

(b) $ka = 0.6$

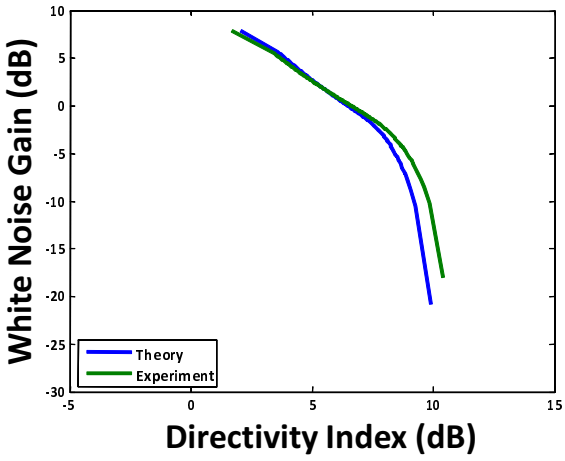


(c) $ka = 1.2$

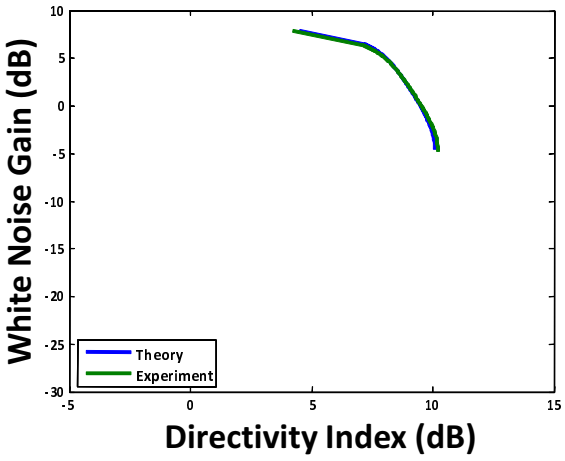


(d) $ka = 2.4$

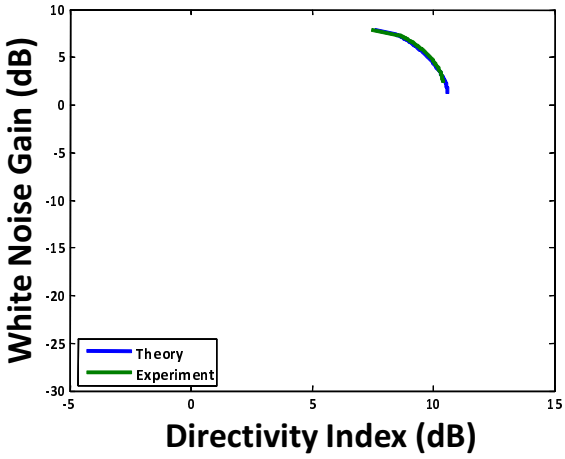
Figure 5.16 Directivity indexes and white noise gains for a 6 microphone spherical array in the free field from theory (blue) and experimental measurement (green) for several values of ka in the 0° azimuth, 0° elevation direction.



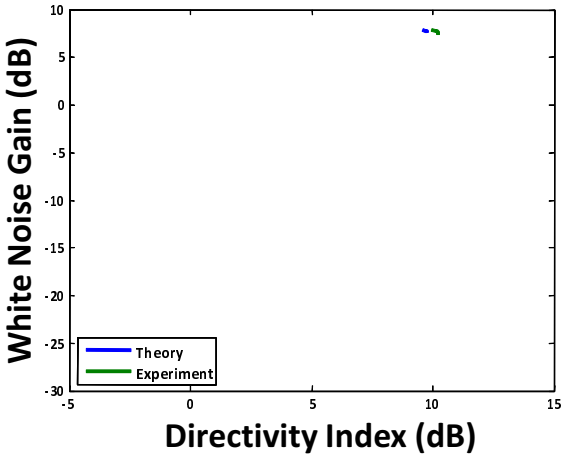
(a) $ka = 0.2$



(b) $ka = 0.6$

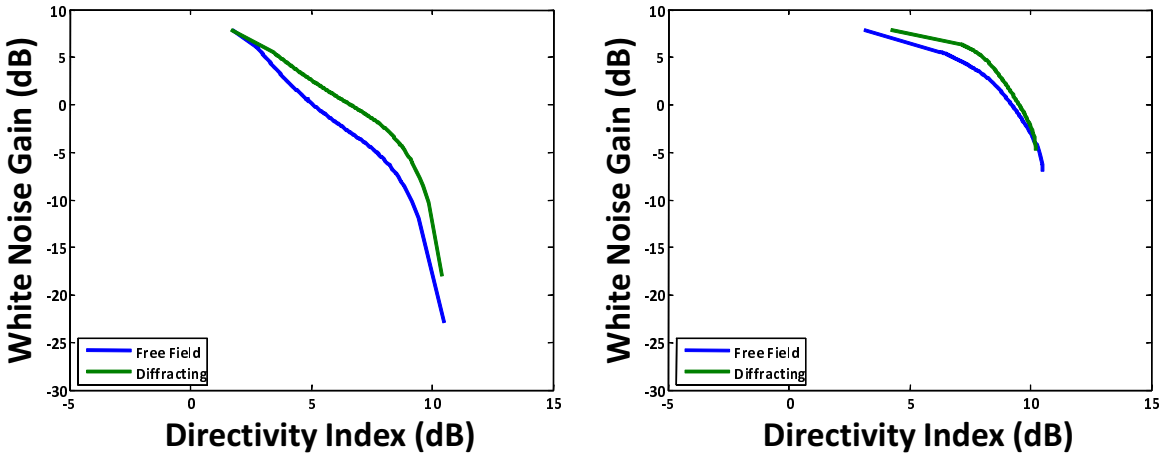


(c) $ka = 1.2$



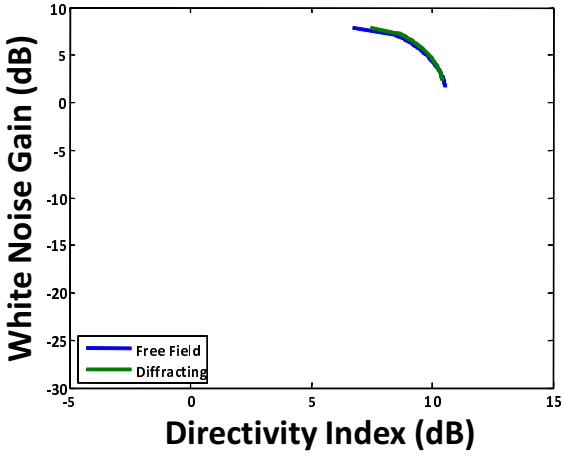
(d) $ka = 2.4$

Figure 5.17 Directivity indexes and white noise gains for a 6 microphone spherical array on a rigid sphere from theory (blue) and experimental measurement (green) for several values of ka in the 0° azimuth, 0° elevation direction.

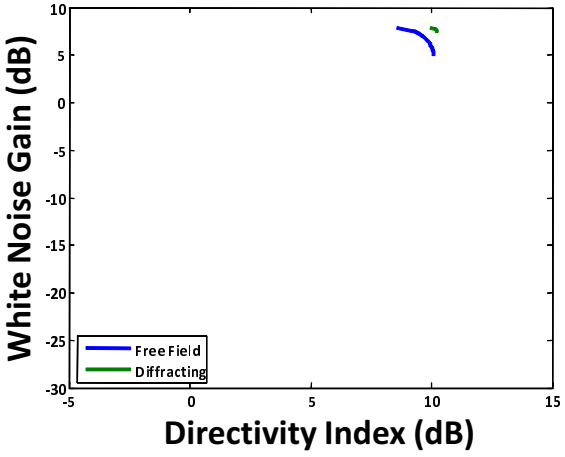


(a) $ka = 0.2$

(b) $ka = 0.6$



(c) $ka = 1.2$



(d) $ka = 2.4$

Figure 5.18 Directivity indexes and white noise gains for a 6 microphone spherical array in the free field (blue) and on a rigid sphere (green) from experiments for several values of ka in the 0° azimuth, 0° elevation direction.

matrix conditioning is such that additional singular values are significant to beamformer performance.

Examining the singular values for both the free field and diffracting spherical arrays will provide an overview of the performance benefit of diffraction for all frequencies. Figure 5.19 shows normalized singular values of the free field and diffracting sphere arrays. As in the circular array case, the singular values for the array mounted on a rigid sphere are greater at any frequency value than for a free field array of identical geometry. Within the scope of the 2 geometries chosen, circular and spherical, the presence of a complementing diffracting body within the array proves beneficial to the performance of array beamformers in terms of suppression of spatially diffuse sources and sensitivity to uncorrelated noise.

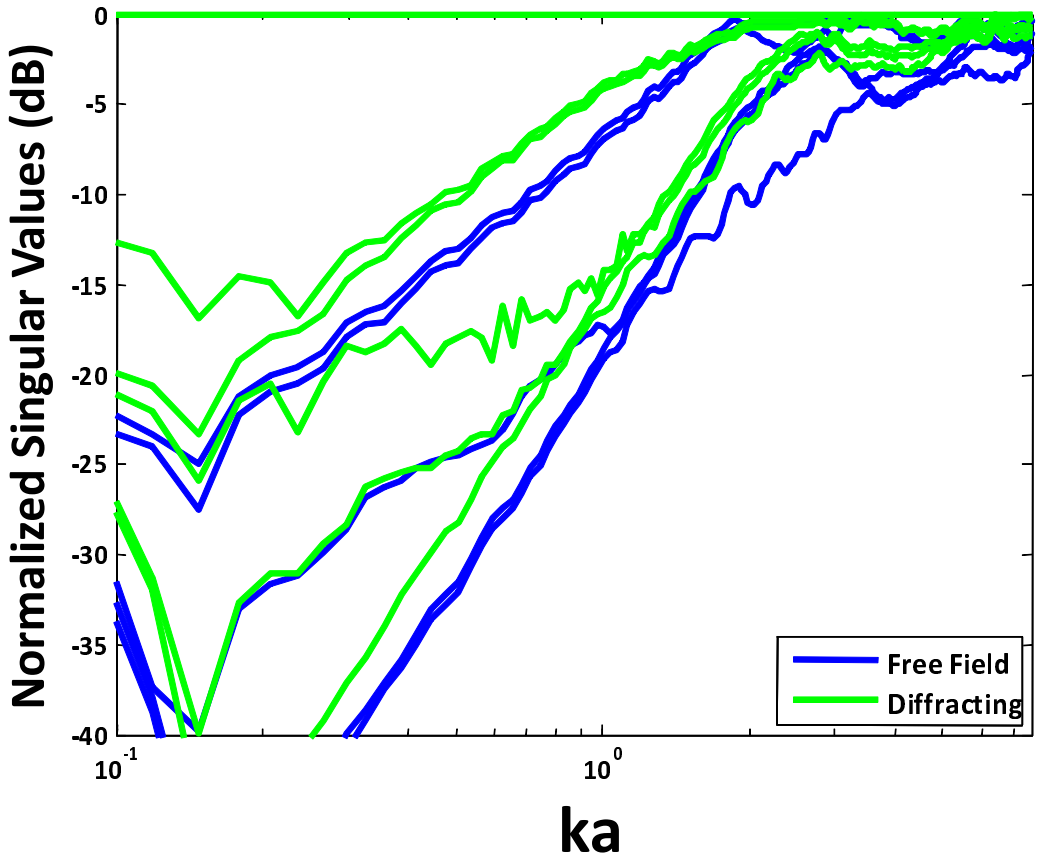


Figure 5.19 Normalized singular values for a spherical array in the free field (blue) and on a diffracting sphere (green).

At lower values of ka the diffracting array is better conditioned and so provides better performance. For example, at $ka = 0.3$ the dipole singular values for the diffracting case are 4dB higher than the free field case. For large values of ka both the free field and diffracting spherical arrays have large singular values leading to good performance.

5.5.3 Helmet

With the underlying advantage of diffracting arrays understood, the placement of an array onto a head mounted platform can be analyzed. The helmet of a military soldier is a logical platform of a microphone array, with individual positions chosen based on the requirements of the system. Embedding microphones onto a helmet, as shown in Figure 5.20, adds negligible weight as supporting systems are likely hand held or otherwise soldier mounted. The capabilities of a helmet mounted system can provide better localization

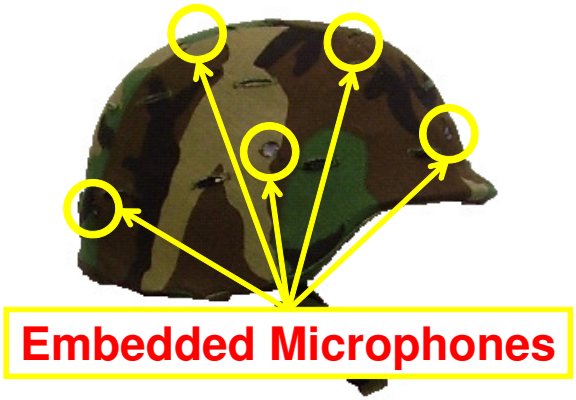


Figure 5.20 Example of helmet mounted microphone array.

and classification ability for sounds in the environment. Whereas there are front-back and up-down ambiguities that may prevent human hearing from accurate localization, these ambiguities and other shortcomings can be overcome through signal processing methods. Additionally, when used in a system equipped with other sensor modalities (image, video, RADAR, etc.) and networking ability, the potential performance increase is substantial [77][78][79]. Given that the performance of an array is highly dependent

upon its geometry, only a cursory comparison of the helmet array will be given. Figure 5.21 shows the singular values for the helmet array compared to the diffracting spherical array. The geometry of the helmet array has 4 microphones placed around the equator of the helmet and 2 microphones equally spaced on the top of the helmet in the median plane of the head (all microphone positions indicated in Figure 5.20 except for one on the far side of the helmet). Notice that although the spherical and helmet array geometries are quite different, their overall performances are comparable.

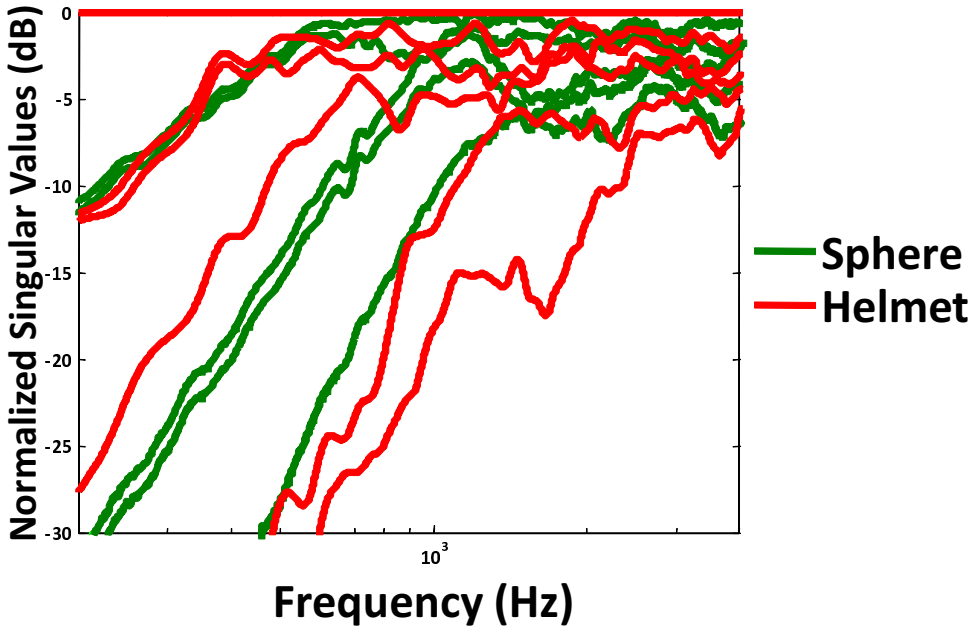


Figure 5.21 Normalized singular values for spherical diffracting array (green) and helmet array (red).

A few implementations of a helmet mounted array are presented in the following sections. The first 2 applications illustrate the capability of the array to track moving tonal objects: helicopters and automobiles. The results presented are meant to show the bearing precision of the array (proving accuracy would require GPS or some other localization system on each vehicle) for various vehicles. The third application presented shows the accuracy of the array in estimating the bearing of supersonic shockwaves from bullets, muzzle blasts from guns, and the combination of those estimations into the localization of the shooter (bearing and range).

5.6 Applications

The helmet array shown in Figure 5.20 was used in several real-world tests to determine the localization performance of the array and processing algorithms. A set of superdirective beamformer (see section 4.1.3) filters were calculated by assuming a cylindrically isotropic noise field. These filters were then used with recorded data to localize sound sources in the environment. Localization was performed by applying filters to blocks of data 1 second long and summing the resulting angular energy distributions. Using gunshot data as an example, Figure 5.22 shows the angular energy distributions for frequencies of 1000Hz, 1500Hz, 2000Hz, and 2500Hz. The summed and interpolated angular distributions for frequencies between 1000Hz and 3000Hz, normalized by the number of frequencies, are shown in Figure 5.23.

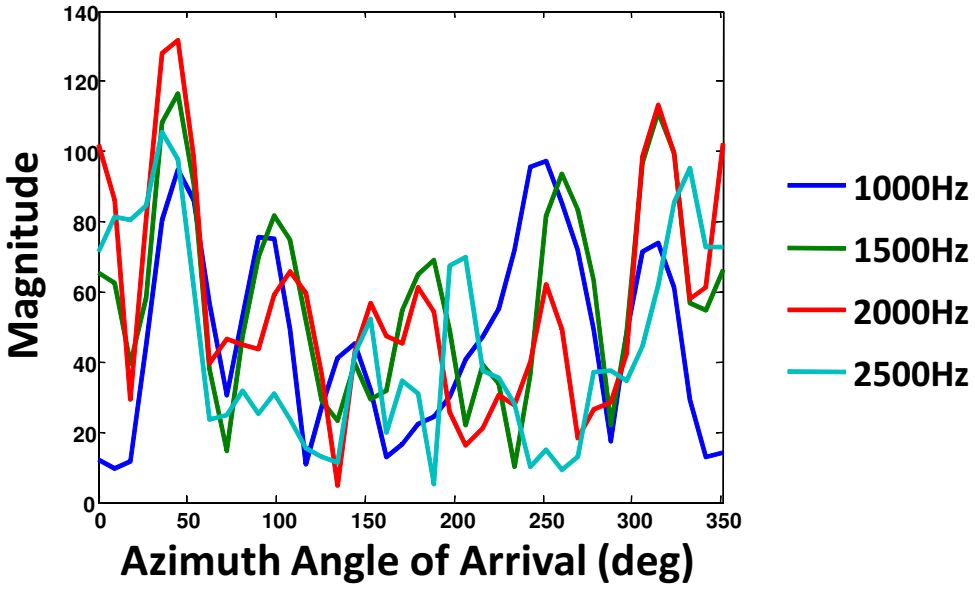


Figure 5.22 Angular energy distributions for four frequencies using gunshot data.

The reason for summing angular distributions is because, in the applications provided, acoustic energy is not concentrated at a single frequency but rather spread over a range of frequencies. For example, in Figure 5.22 the four distributions provided appear to indicate a source between 36° and 45°, although a source at 315° may seem just as likely. However, once all distributions for frequencies between 1000Hz and 3000Hz are summed

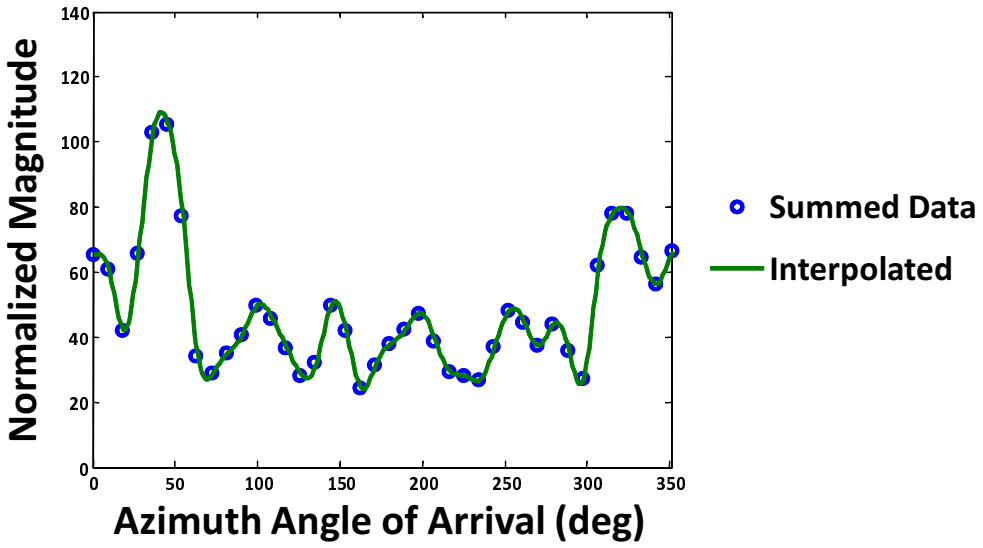


Figure 5.23 Summed angular energy distribution and interpolated angular energy distribution for a frequency range of 1000Hz to 3000Hz.

together and normalized, the resulting distribution quite clearly shows that the source is indeed between 36° and 45°, as seen in Figure 5.23. Using interpolation on the summed distribution (assuming the individual distributions are simply digitizations of the true distribution) provides an estimate of 42°, also shown in Figure 5.23.

The following tests involved the localization of AH-64 Apache helicopters at an airport, automobiles on a road, and gunshots at a shooting range.

5.6.1 Helicopter Localization

As is generally the case for Virginia Tech home football games, a stadium flyover was planned using AH-64 Apache helicopters. The flyover was to involve 3 helicopters: 2 in the actual flyover and 1 on the runway at the Montgomery County Executive Airport during the flyover. The helmet array was placed on a human who stood approximately 100 yards from the helicopters, as shown in Figure 5.24. For the simplicity of the test explanation, the helicopters will be numbered 1 to 3 from left to right in Figure 5.24.

During a minute of measurement several events occurred, as shown in Figure 5.25:

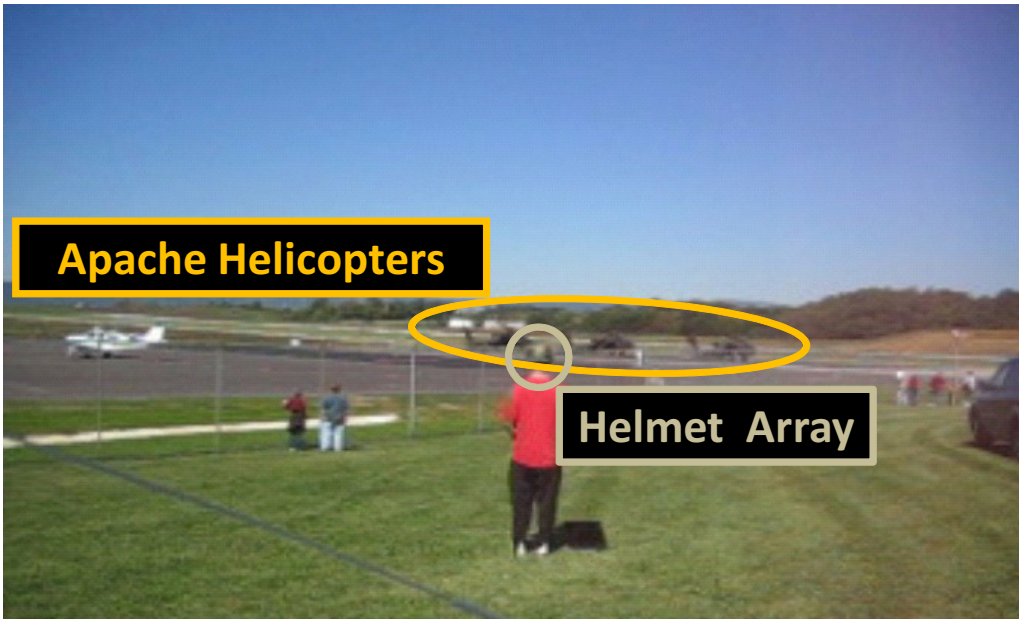
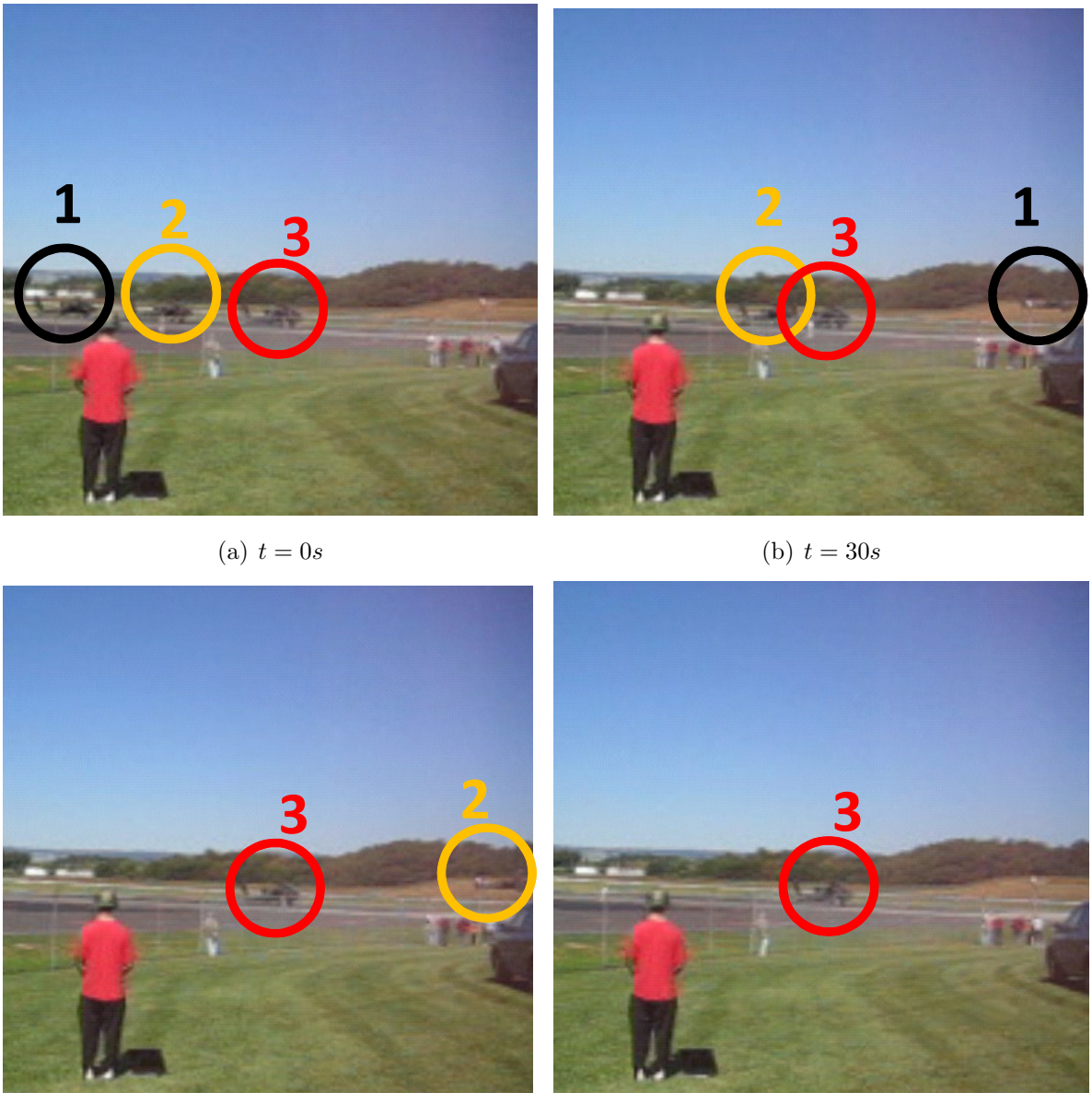


Figure 5.24 Helmet array (gray) worn by human standing approximately 100 yards from 3 AH-64 Apache helicopters (orange) at the Montgomery County Executive Airport.

- Helicopter 1 powered up for taxiing while helicopters 2 and 3 idled on the tarmac (Figure 5.25(a))
- Helicopter 1 taxied partially down the runway before taking off and hovering the remainder of the way (Figure 5.25(b))
- Helicopter 2 powered up, taxied partially, took off, and hovered the remainder of the way (Figure 5.25(c))
- Helicopter 3 powered up in place (Figure 5.25(d))

The acoustic information recorded during the events in Figure 5.25 is shown in Figure 5.26(a) while the results from bearing estimation are shown in Figure 5.26(b). Since the helicopters all had roughly the same acoustic signature, the presence or absence of an acoustic event was determined by the amount of energy present in the measured signals. For example, as helicopter 1 powers up, helicopters 2 and 3 are idling on the tarmac. The amount of acoustic energy present from helicopter 1 is greater than from helicopters 2 and/or 3, and the bearing estimation favors the direction of helicopter 1. The bearing



(a) $t = 0s$

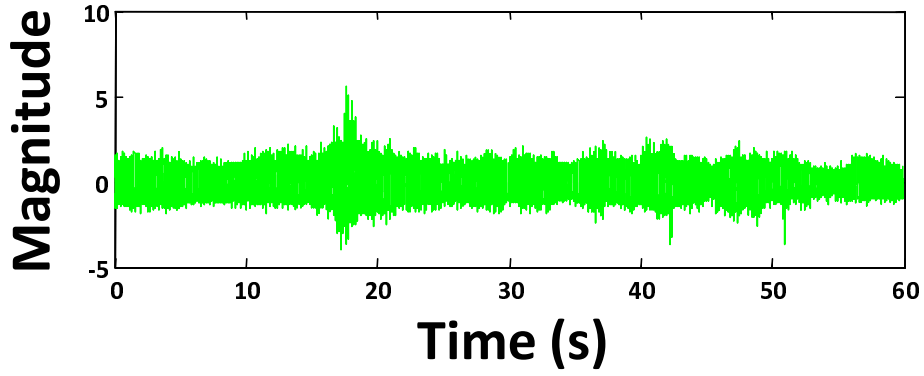
(b) $t = 30s$

(c) $t = 43s$

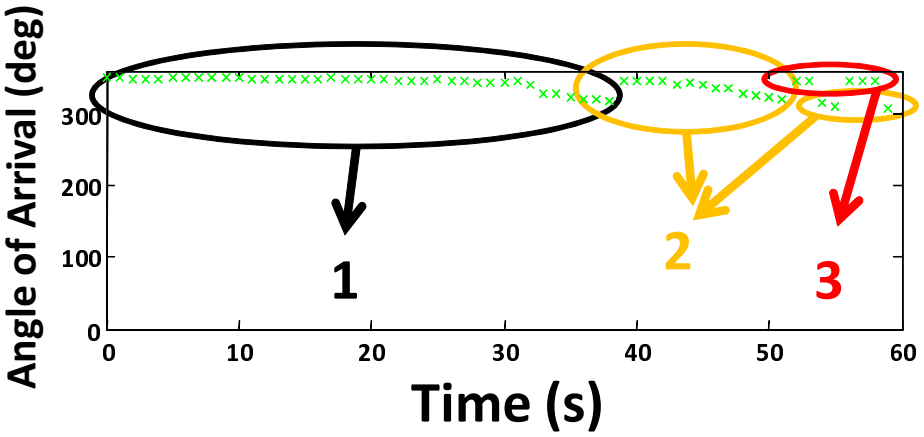
(d) $t = 55s$

Figure 5.25 Acoustic events occurring at stated times during one minute test. Helicopters are highlighted and labeled.

track from helicopter 1 is very consistent and follows the position of the helicopter until it passes behind the airport terminal (to the right of the frame in all Figures).



(a) Acoustic data from microphone 1



(b) Bearing estimation

Figure 5.26 Acoustic data from microphone 1 of the helmet array and bearing estimation of acoustic events during AH-64 Apache helicopter test.

One caveat is that helicopter 3 begins to power up while helicopter 2 continues to taxi and hover. At that point the acoustic energy present from each helicopter is nearly equal, producing a bit of uncertainty in the bearing estimation (last 10 seconds in Figure 5.26). A simple conclusion is that this is one of the more difficult cases to localize, i.e. one helicopter in the midst of many similar helicopters. That is to say that if a group of Apache attack helicopters is about to attack, the least worrisome issue will be whether the system is localizing the helicopter in the front left of the pack or the front right.

5.6.2 Automobile Localization

Another example of the helmet sound localization is illustrated using vehicles driving by on a road. The helmet array was placed on a post approximately 5 feet higher than the road and 10 feet from the road, simulating the height of a human head, and array measurements were taken for post-processing. A view of the experiment setup is shown in Figure 5.27. The cars that passed during the test will be numbered in chronological order, 1 through 6.

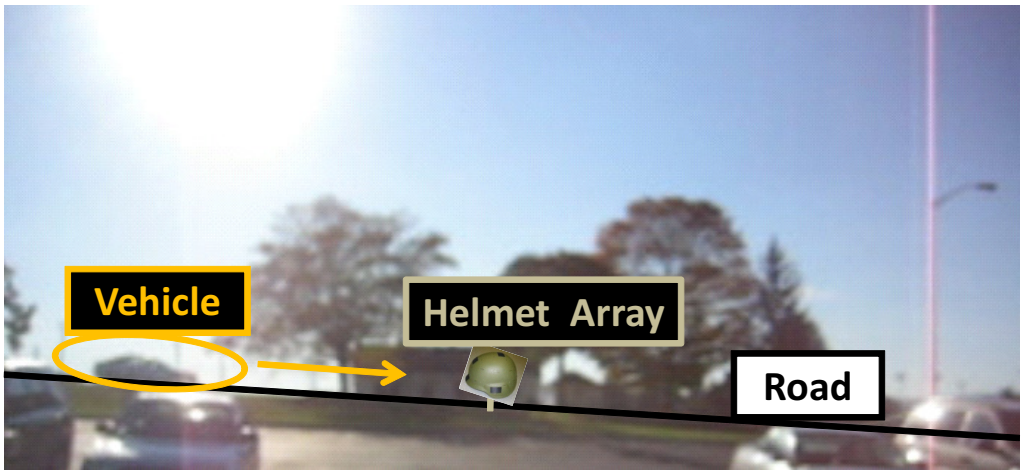


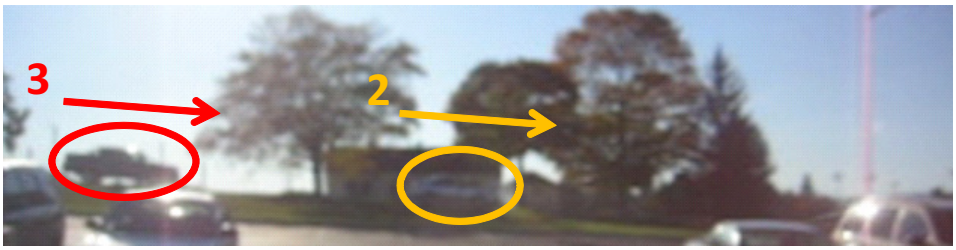
Figure 5.27 Helmet array (gray) placed next to a busy road (black) with vehicles driving by from both sides (orange).

During 30 seconds of measurement several events occurred, as shown in Figure 5.28:

- Automobile 1 approaches from the left and departs to the right (Figure 5.28(a))
- Automobiles 2 and 3 approach from the left and depart to the right (Figure 5.28(b))
- Automobile 4 approaches from the right and departs to the left (Figure 5.28(c))
- Automobiles 5 and 6 approach from the left and depart to the right (Figure 5.28(d))

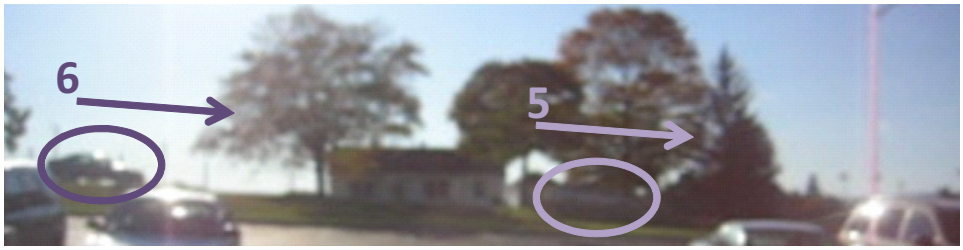
The acoustic information recorded during the automobile tracking test is shown in Figure 5.29(a) while the results from bearing estimation are shown in Figure 5.29(b). The presence or absence of an acoustic event was again determined by the amount of energy

(a) $t = 1s$



(b) $t = 5s$

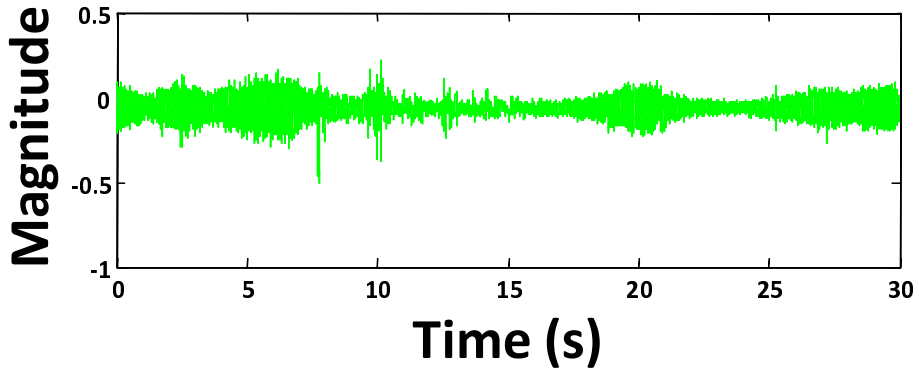
(c) $t = 21s$



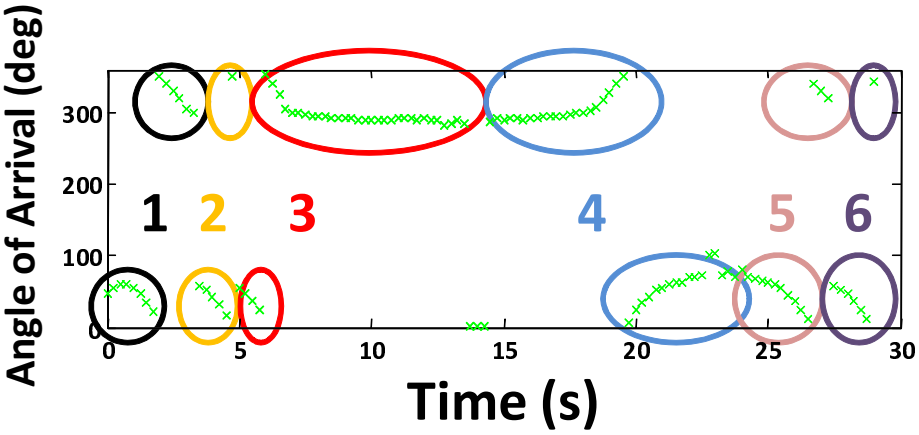
(d) $t = 28s$

Figure 5.28 Acoustic events occurring at stated times during 30 second automobile tracking test. Automobiles are highlighted and labeled.

present in the measured signals. For example, as soon as automobile 2 is past the array, automobile 3 immediately passes the array. Given the closer proximity to automobile 3 as compared to 2, more acoustic energy measured by the array comes from automobile 3. This pattern is visible within the transitions between all automobiles.



(a) Acoustic data from microphone 1



(b) Bearing estimation

Figure 5.29 Acoustic data from microphone 1 of the helmet array and bearing estimation of acoustic events during automobile tracking test.

5.6.3 Gunshot Localization

Research into the nature of supersonic gunshots has shown that the acoustic information present in the passing bullet shockwave and the gun muzzle blast is enough to

estimate an exact shooter location. Although the acoustic localization presented is passive in nature, the 2 aforementioned pieces of information about a single gunshot provide the ability to triangulate the gunshot location. Consider a gunshot that occurs a distance d and angle θ_b from a microphone array, as shown in Figure 5.30. The shockwave from the passing bullet is formed as a conical wave from the front of the bullet. The angle of the triangular wavefront, θ_{shock} , is dependent upon the Mach number, M, of the bullet,

$$\theta_{shock} = \arcsin \frac{1}{M} \tag{5.11}$$

In contrast, the muzzle blast from the gunshot spherically spreads from the end of the gun barrel. Assume the bullet shockwave arrives at the array at time t_1 from angle θ_s. The muzzle blast acoustic wave will arrive at some later time t_2 from angle θ_b. The differences in time $\Delta t = t_2 - t_1$ and angle $\theta_d = \theta_b - \theta_s$ are used to calculate the distance from the array to the location of the gunshot,

$$d = \frac{\Delta t c}{1 - \cos \theta_d} \tag{5.12}$$

where c is the speed of sound in air based on atmospheric conditions (e.g. temperature, humidity, barometric pressure, etc.).

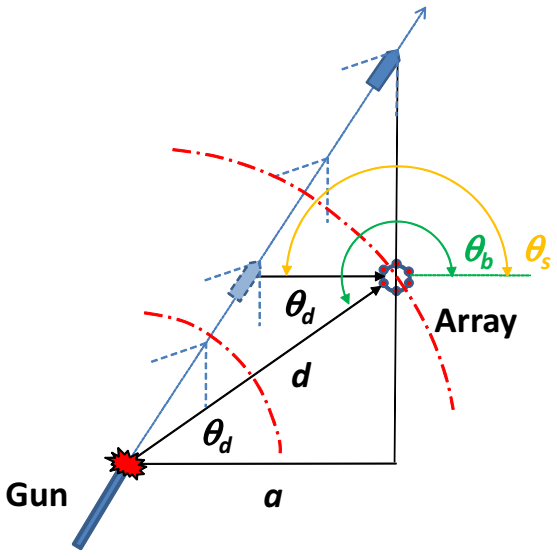


Figure 5.30 Geometry for localization of supersonic gunshot.

A preliminary test performed with 2 linear acoustic arrays provides a useful example. Figure 5.31 shows the experiment setup at the shotgun range near Virginia Tech. The sound measured from 3 microphones on 1 line array are shown in Figure 5.32. The first set of impulses are from the supersonic shockwave as the bullet passes the array, shown in Figure 5.33(a), and the second set of impulses are from the muzzle blast of the initial gunshot, shown in Figure 5.33(b).

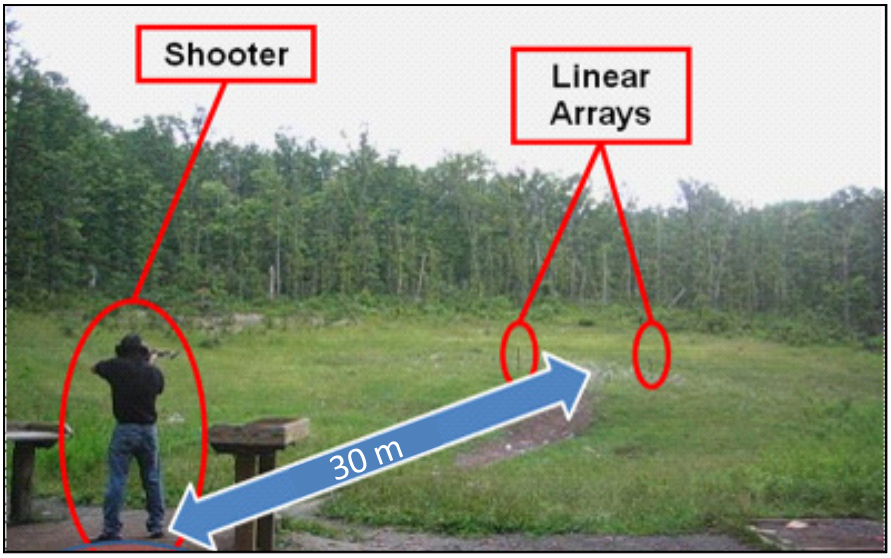


Figure 5.31 Shooting range setup for supersonic bullet localization experiment.

In the line array example the shooter was approximately 30 meters from the array. The measured time difference was $\Delta t = 0.0445s$ and the estimated angle difference was $\theta_d = 54.8°$, producing an estimated range of $d = 36.3m$, or approximately 20% higher than the true value. The helmet array was used in a similar test using 3 different rifles. An AR-15 rifle, the civilian version of the military M-16, an SKS rifle, using the same round as the AK-47, and a .22-magnum rifle, a high velocity version of a .22 rifle, were used to test the method on bullets with different supersonic speeds. The results of the test are shown in Table 5.1 for 15 shots taken at a range of 56 meters. One shot from the .22-magnum was not localized due to the absence of a supersonic shockwave (at 56 meters the bullet has slowed significantly, to a point near, and in this case, below the speed of sound). There is some disparity between the detected angles for each of the different

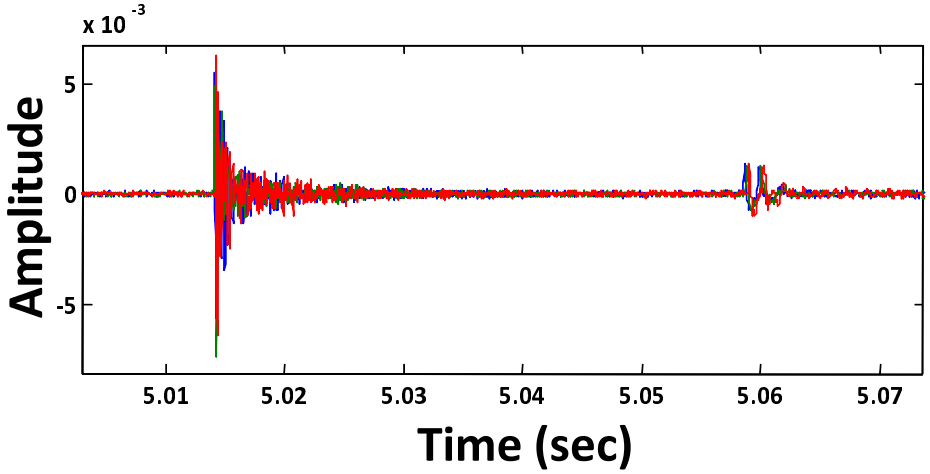
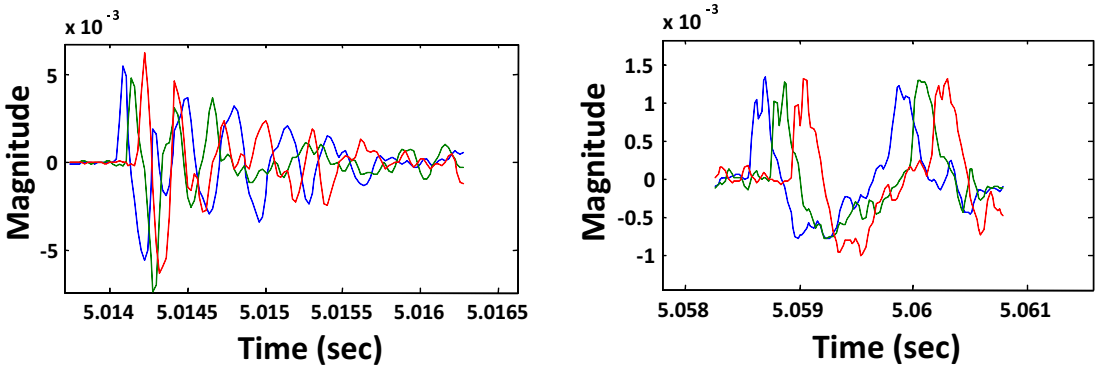


Figure 5.32 Sound measured from 3 microphones on 1 line array during supersonic bullet localization experiment.



(a) Shockwave

(b) Muzzle blast

Figure 5.33 Impulses from supersonic bullet shockwave and muzzle blast.

Table 5.1 Bearing and ranging results from supersonic gunshot testing.

Firearm	Detected Angle	Detected Range	Range Error	% Range Error
AR-15	77	70	14	25%
AR-15	78	61	5	9%
AR-15	76	67	11	20%
AR-15	76	68	12	21%
AR-15	77	67	11	20%
SKS	82	67	11	20%
SKS	82	68	12	21%
SKS	84	63	7	13%
SKS	81	61	5	9%
SKS	82	67	11	20%
.22M	91	76	20	36%
.22M	91	71	15	27%
.22M	91	77	21	38%
.22M	91	68	12	21%

rifles. This disparity was likely due to the high winds and an "accident" changing the orientation of the manikin on which the helmet was mounted. A set of gunshots for each gun took less than 30 seconds while the changeover time between guns was on the order of several minutes, giving ample time for wind to influence the array orientation. The "accident" occurred between the SKS and .22-magnum rifles when the manikin was blown over and broken in half, requiring a reset of the manikin torso and helmet into its original position (performed by sight). The tight angle groupings for each gun fired (largest disparity is 3° for the SKS) indicates that the disparity in angle was likely due to an absolute orientation uncertainty and not the localization method or helmet array.

CHAPTER 6

Voice Isolation

6.1 GSC Applied to a Headset Voice Isolation System

In this chapter the results of a voice isolation system will be presented. A voice isolation system is one where the objective is to extract the user's voice signal while operating in a noisy environment. Such a system can be used to enhance the performance of boom microphones or to remove the need for a boom entirely. The system presented here will consist of a six microphone array mounted on a headset (i.e. 3 mics on each earcup) and will demonstrate the potential of such a system especially when the noise signals are tonal in nature. Potential applications include cell phone headsets that are barely the size of the human ear and are required to isolate the speech of the user while suppressing noise from all other directions [80]. Current hearing aid technology offers users settings for different noise environments such as close conversations in noisy crowds [81]. These applications illustrate the inherent mix between near-field and far-field sources that exist in real world environments. In the near-field the array manifold vector for the direction of the mouth can be treated similarly to the vectors for far-field sources. However, the relative magnitude and phase differences in the near-field vary much more rapidly with changes in position than in the far-field, possibly causing signal mismatch errors if slight changes in array or microphone positions are experienced.

Fundamentally, the voice isolation method described here (because there are many methods available) involves two separate beamformers: a "beam-to-voice" filter and a "block-voice" filter. All beamformer filters are calculated *a priori*, i.e. they are data independent filters, operating under the assumption of a fixed head-array geometry relationship. The microphone signals are filtered by each filter to obtain a "mostly voice" signal and a "without voice" signal. The idea is that the "mostly voice" signal may still have spatial interferer's while the "without voice" signal should have only spatial interferer's since in principle it is possible to perfectly notch out a signal in a known direction i.e. the voice. The "without voice" signal is then used as a noise reference signal to filter out all coherent information from the "mostly voice" signal to arrive at an optimally beamformed signal. The implementation uses 2 data independent beamformers, one beaming to the look direction and one beaming to everything but the look direction, and an adaptive process to optimally filter the data is known as a linearly constrained minimum variance beamformer (LCMV), or more specifically, a generalized sidelobe canceller (GSC) [82]. Figure 6.1 shows a block diagram of the filtering process used to isolate the voice signal.

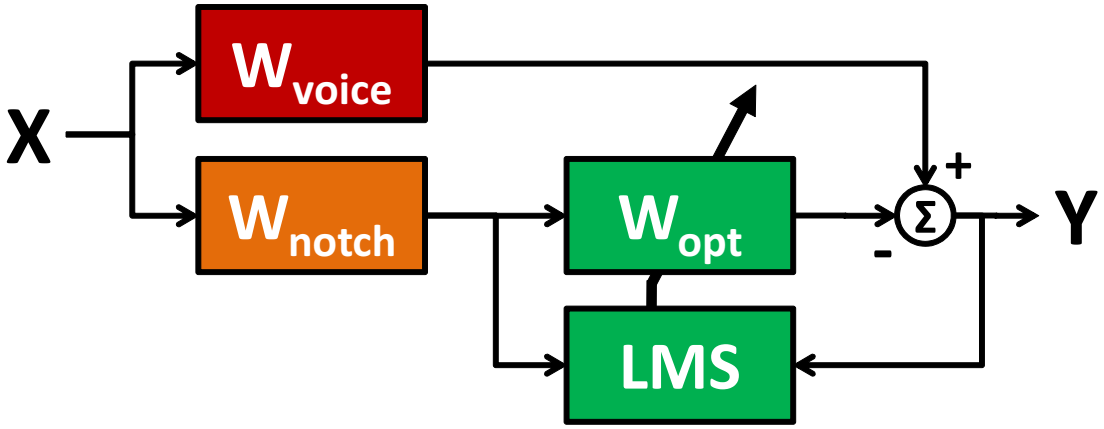
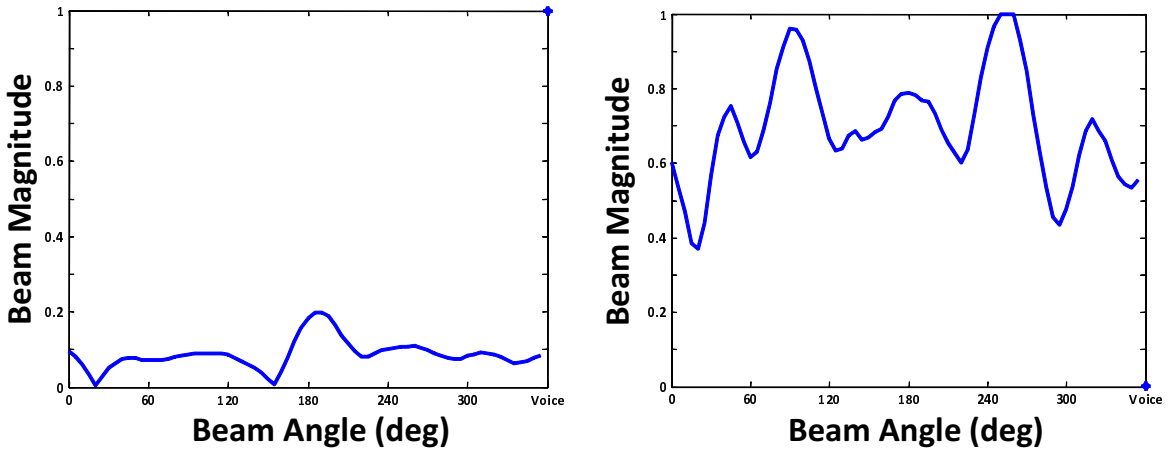


Figure 6.1 Voice isolation filtering block diagram.

The calculation of "beam-to-voice" and "block-voice" filters is accomplished through the method described in Equations 4.22, 4.23, 4.24, and 4.25. An example of the "beam-to-voice" filter beampattern is shown in Figure 6.2(a) and a "block-voice" filter beam-

pattern is shown Figure 6.2(b). Within Figure 6.2, there are azimuth angles between 0° and 360° and the voice is then indicated by an asterisk on the right side of the graph. It should therefore be noted that the sensitivity to the voice is high for the "beam-to-voice" beampattern and effectively zero for the "block-voice" beampattern. Placing an interferer at an azimuth angle of 250° and applying the GSC yields the beampattern shown in Figure 6.3.



(a) Beam-to-voice

(b) Block-voice

Figure 6.2 Beampatterns for "beam-to-voice" and "block-voice" filters for an 8 microphone headset array for a frequency of 1500Hz.

6.2 Headset Array Experimental Results

The aforementioned voice isolation filtering methods were demonstrated using a headset mounted microphone array placed on a manikin. Using the characterization method described in Chapter 3, array manifold vectors were measured for azimuth directions at 0° elevation. Additionally, a speaker mounted in the location of the manikin's mouth, shown in Figure 6.4, was used to measure the array manifold vector for the voice direction. All array manifold vectors were then combined to form a single array manifold matrix for the headset array. Recorded signals were processed with a set of filters calcu-

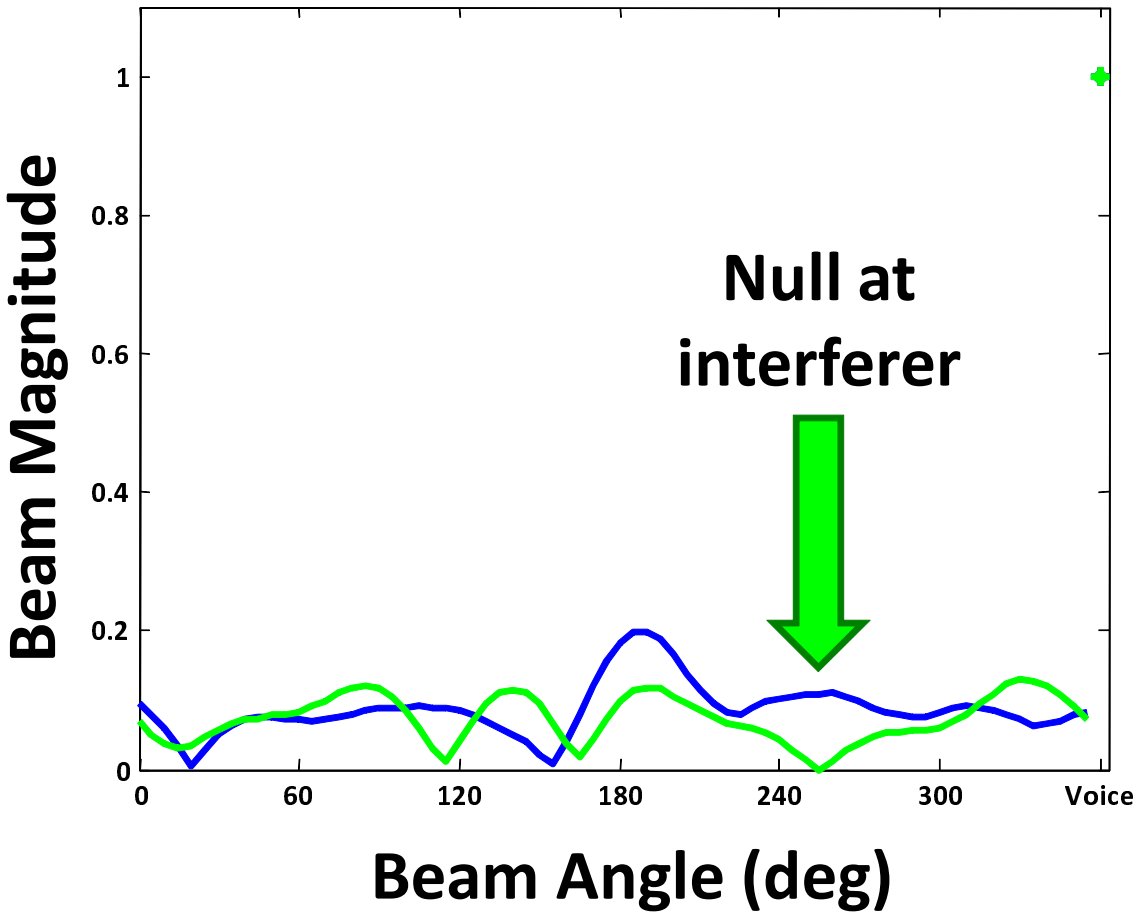


Figure 6.3 Adapted voice isolation beampattern (green) after application of GSC method for an 8 microphone headset array and an interferer at 250° at a frequency of 1500Hz, compared to the "beam-to-voice" beampattern (blue).

lated from the manifold measurements and then used with the LMS algorithm described in section 4.1.4.1.

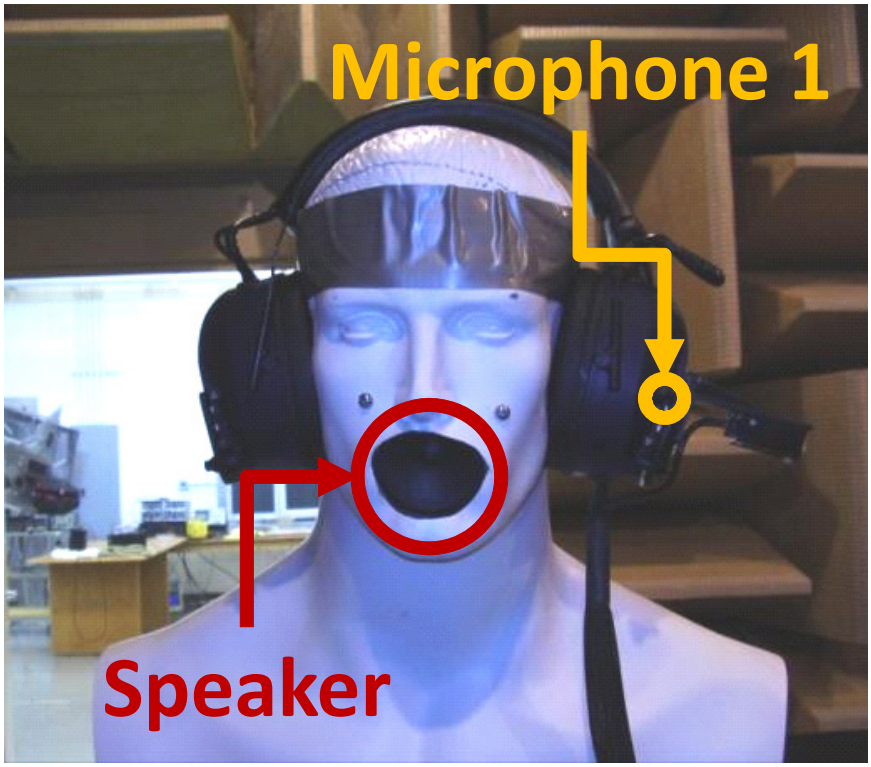


Figure 6.4 Manikin wearing headset array with speaker mounted in mouth(red).

The first experiment was setup using the speaker in the manikin's mouth and an external speaker approximately 6 feet away at 90° relative to the manikin (directly to the left of the manikin), as shown in Figure 6.5. A second experiment was performed with the external speaker the same distance away but at 0° relative to the manikin (directly in front of the manikin), as shown in Figure 6.6. The mouth speaker played a pre-recorded human voice reading a newspaper while the external speaker played either narrowband or broadband noise.

There are a few compromises made to analyze the experimental results in the following sections. The only signals recorded were from the headset array, i.e. no external signal or noise references were used to augment the performance of the voice isolation algorithm. However, the lack of reference signals complicates the presentation of signal-to-noise ratio

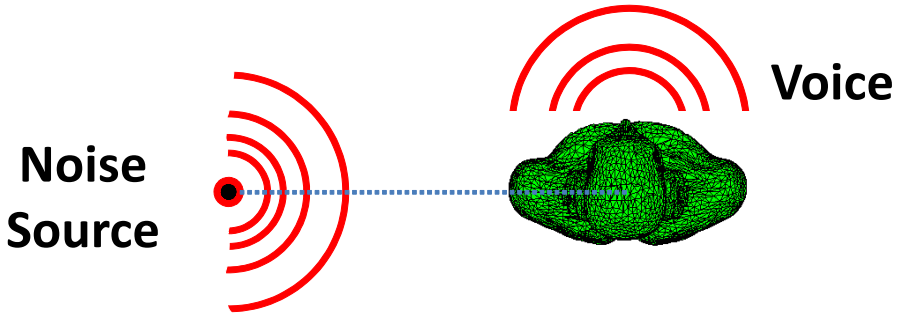


Figure 6.5 First experiment setup, with interferer at 90°, for voice isolation algorithm testing.

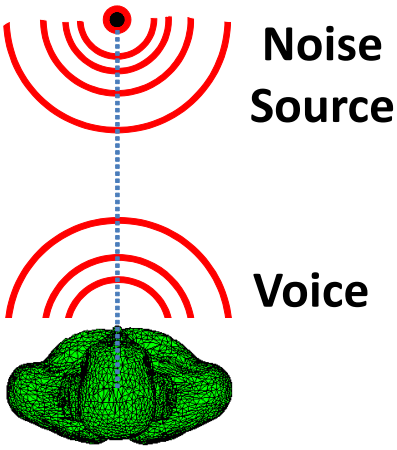


Figure 6.6 Second experiment setup, with interferer at 0°, for voice isolation algorithm testing.

(SNR) improvements because no pure "signal" or "noise" was available for comparison. Therefore, in lieu of presenting a conventional SNR improvement for each case, the difference between filtered and unfiltered signals will be shown in the frequency domain to show the magnitude of noise suppression.

One caveat about the LMS algorithm in this application is that it works best when the SNR is very low (i.e. the noise is completely drowning out the signal). Since there is no pure reference "noise" signal to remove, the reference signal is determined by "beaming to" or "notching out" particular directions. Any residual signal present in the "beamed reference" will then be applied in the LMS algorithm as something to remove. Therefore, if the SNR is high (i.e. the desired signal is much louder than the noise), "beaming to" the noise will likely produce a reference signal that is part noise and part signal. When that reference signal is used to adaptively filter the "beam-to" signal, it will act upon the signal just as it will the noise. A counter to this caveat is: If the SNR is high to begin with, adaptive noise filtering is probably not necessary. For intermittent cases cases, such as speech (i.e. where the signal is not continuously available but rather starts and stops with syllables), the LMS can be applied only when the energy in the "beam-to-voice" signal is above a threshold.

Another compromise made is the comparison between raw microphone signals and the output of the adaptive algorithm. While there were 6 microphones in the array there was only 1 output from the algorithm, complicating the decision of which microphone to compare to the filtered results. The following results present the voice isolation algorithm results versus the raw signal from microphone 1 in the array, showing the improvement provided over a single microphone case. The position of microphone 1, on the left headphone mounted in a frontward orientation, is highlighted in Figure 6.4.

6.2.1 Narrowband Noise

6.2.1.1 Interferer at 90°

The first noise source used as an interferer was comprised of a series of tones starting at 200Hz, spaced 200Hz apart, and ascending up to 4000Hz. The results from this experimental setup are shown in Figure 6.7 for the raw signal from a microphone (blue), the beamformed signal output using the entire array (green), and the output of the voice isolation (or GSC) algorithm (red). The amount of noise suppression for the 2 filtered cases is shown in Figure 6.8.

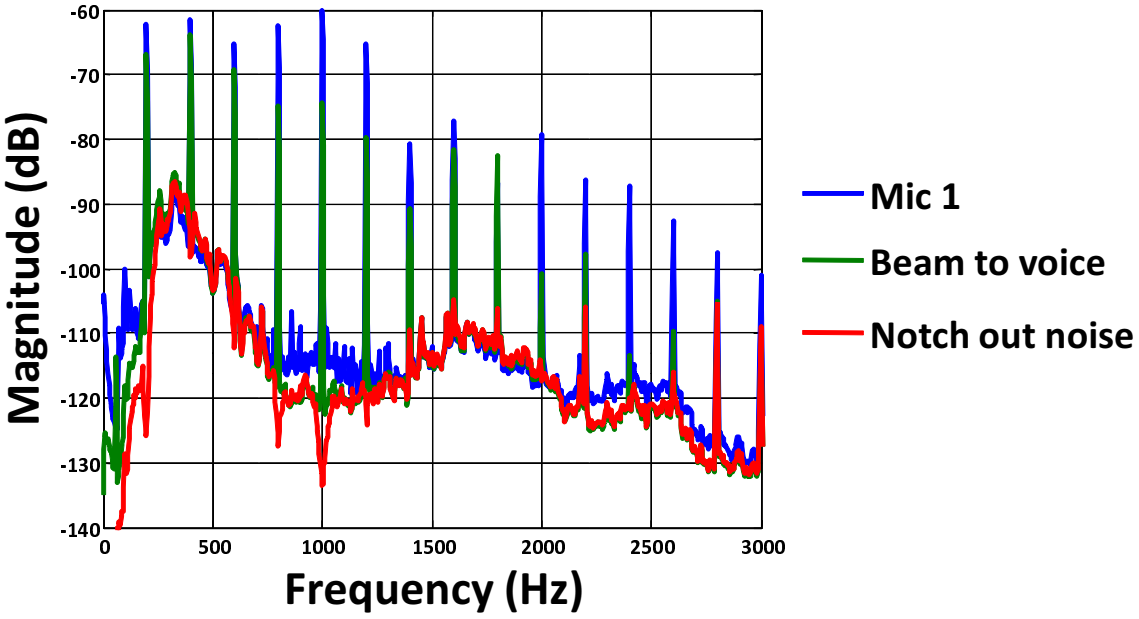


Figure 6.7 First experiment setup results of voice isolation test with an interfering tones signal for a single microphone signal (blue), beamformed output (green), and voice isolation algorithm output (red).

Within Figure 6.7, the first feature to notice is the dominance of the interfering tones over the voice signal. The tones are, at best, 60dB above the level of the voice signal and, at worst, nearly 130dB above. The output from the "beam-to-voice" beamformer shows some improvements over the raw microphone signal; in most cases a reduction of 30dB. However, the combination of the "beam-to-voice" and the "block-voice" beamformers in the GSC yield great results. The tones are all but eliminated for every frequency below

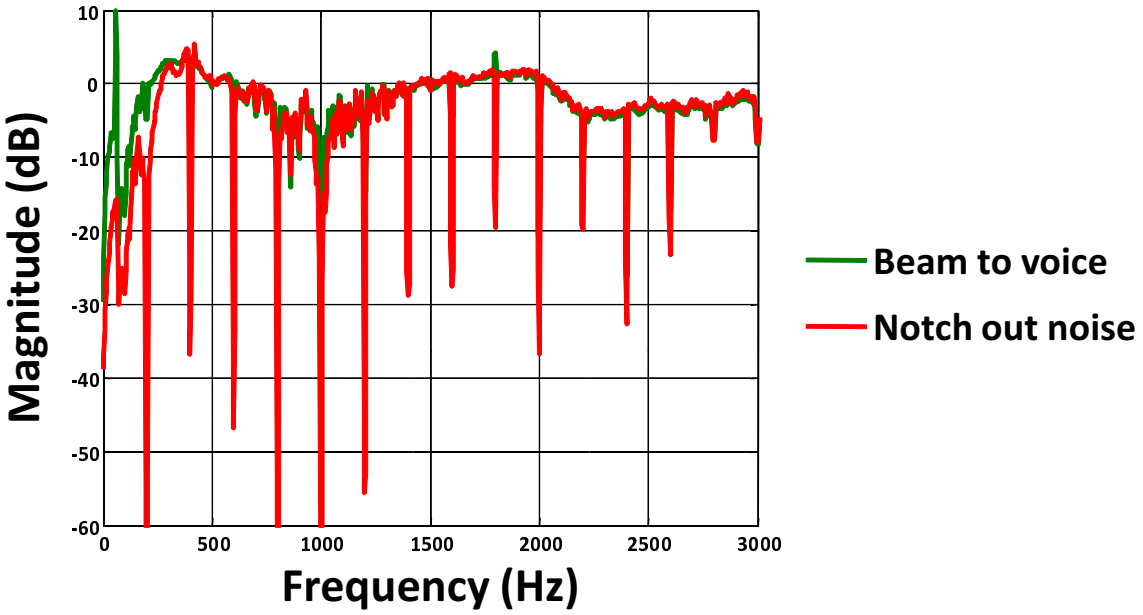


Figure 6.8 Noise suppression improvement over single microphone case for interfering tones at 90° for beamformed output (green), and voice isolation algorithm output (red).

approximately 2800Hz and the frequency domain content of the voice is visible below and near 500Hz (the pre-recorded voice was male). Note in Figure 6.8 how the "beam-to-voice" beamformer is able to provide up to 20dB of noise reduction between 500Hz and 1000Hz, but overall drastically reduce all of the tones. In contrast, the addition of the "block-voice" beamformer provides a minimum reduction of 7dB for tones at 2800Hz and 3000Hz and massive 60dB reductions at 200Hz, 800Hz, and 1000Hz. The additional performance gained by adaptively removing the interfering signal from the "beam-to-voice" signal is undeniable.

6.2.1.2 Interferer at 0°

Using the same noise source as the previous section, the external speaker was moved directly in front of the manikin for this second test. The goal of this test was to establish the ability of the array and GSC algorithms to discern between nearfield and far-field sources in the same direction. Very often the far-field assumption is used to make array manifold matrices easier to calculate, as all sources in a particular direction are treated

as having plane wave incidence. In this case, the fact that the voice is in the near-field means that the array manifold vectors for the external speaker direction and the voice direction will likely be different. This difference provides an avenue for the GSC algorithm to isolate the voice and minimize the interfering signal from the external speaker. The results of this second test are shown in Figure 6.9 with a single microphone signal (blue) and the output of the voice isolation algorithm (red). The amount of noise suppression for the 2 filtered cases is shown in Figure 6.10.

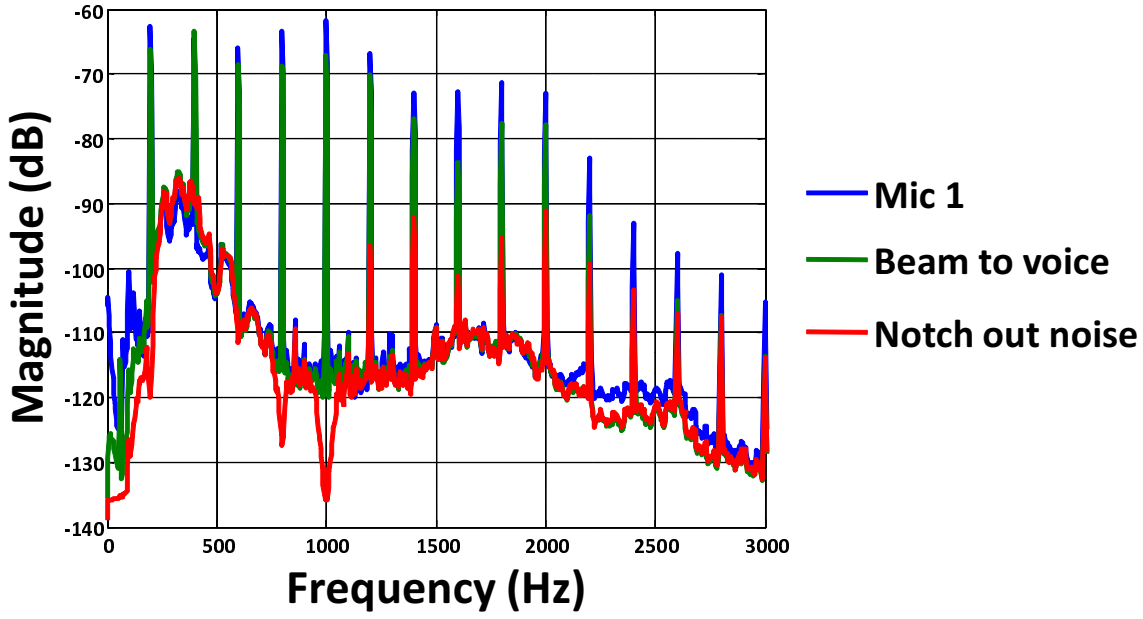


Figure 6.9 Second experiment setup results of voice isolation test with an interfering tones signal for a single microphone signal (blue) and voice isolation algorithm output (red).

As in the case for the interferer at 90°, the performance of the voice isolation filtering is impressive. The "beam-to-voice" filter provides reductions of 5dB for the 800Hz and 1000Hz tones and modest improvements otherwise. However, adding the adaptive "block-voice" filtering again produces remarkable results. Tones at 2800Hz and 3000Hz are still reduced by 7dB while ALL interfering tones at and below 1200Hz are reduced by at least 35dB. The improvement in the "block-voice" case over the "beam-to-voice" case shows the ability of the LMS algorithm to exploit the differences between the magnitude and

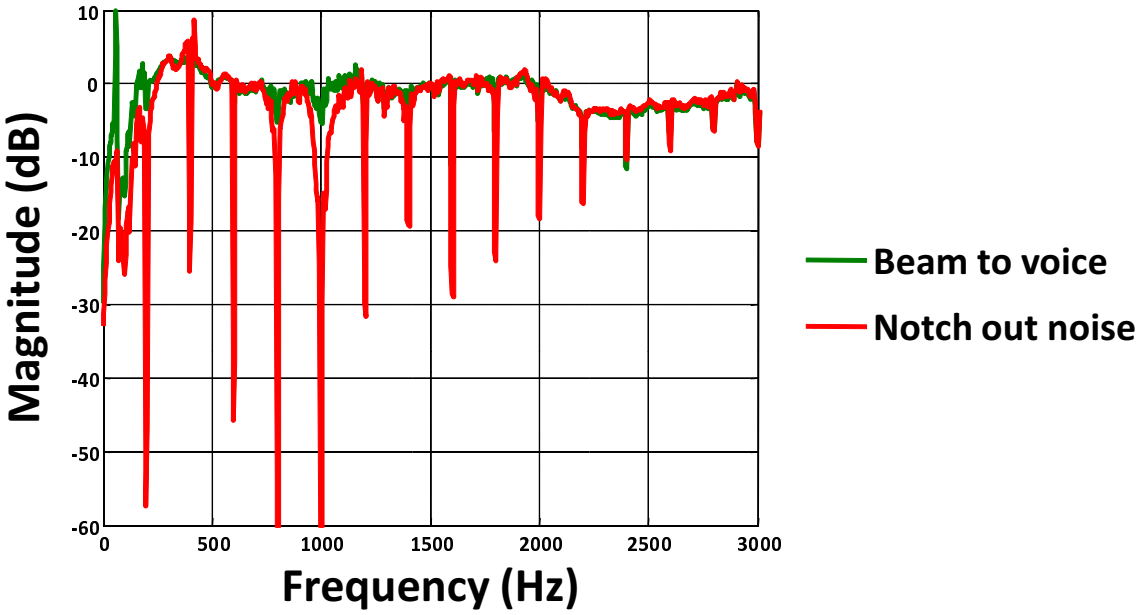


Figure 6.10 Noise suppression improvement over single microphone case for interfering tones at 0° for beamformed output (green), and voice isolation algorithm output (red).

phase of the near-field voice signal and the far-field external speaker signal. This also illustrates the degree to which a particular direction can be "beamed-to" versus how well a direction can be "notched-out." Beaming to the voice provided modest gains over the single microphone case but notching the voice and using the resulting signal as a noise reference for adaptive noise suppression provided great results.

6.2.2 Broadband Noise

The second noise source used as an interferer was broadband noise between 20Hz and 4000Hz. As in the previous case the tests were conducted using an interferer at 90° and then at 0°. While the interfering tones added energy to a few frequencies throughout the voice frequency range, the interfering broadband noise added energy to all frequencies in that range.

6.2.2.1 Interferer at 90°

The results from the first broadband experiment setup (voice signal from manikin's mouth speaker and broadband interferer from external speaker placed at 90°) are shown in Figure 6.11. The amount of noise suppression for the 2 filtered cases is shown in Figure 6.12.

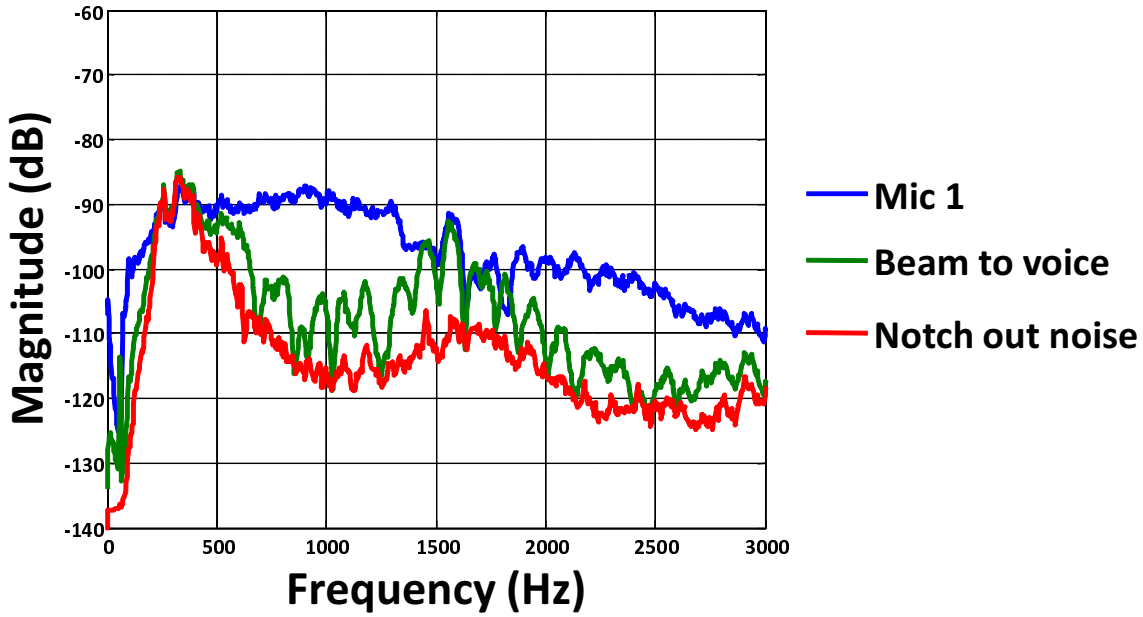


Figure 6.11 First experiment setup results of voice isolation test with an interfering broadband signal for a single microphone signal (blue), beamformed output (green), and voice isolation algorithm output (red).

In the raw microphone signal (blue) the noise floor clearly drowned out any semblance of voice signals. The "beam-to-voice" beamformer (green) attenuated the broadband noise to a degree that revealed some of the underlying voice signal. Application of the GSC algorithm (red) clearly outperformed the "beam-to-voice" beamformer, suppressing the interfering broadband noise and leaving the voice signal intact. From Figure 6.12 the "beam-to-voice" beamformer appeared to work very well at some frequencies (30dB suppression just above 1000Hz) but inconsistently across the frequency range (15dB suppression just below 1000Hz). Addition of the "block-voice" beamformer improved upon the performance of the "beam-to-voice" beamformer, specifically where the inconsistent

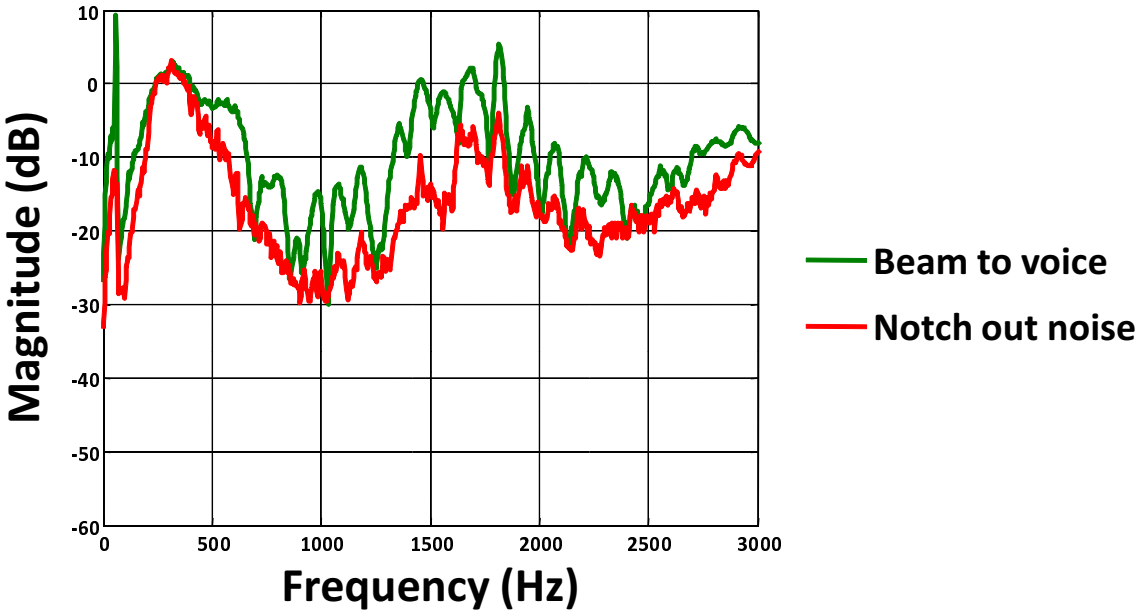


Figure 6.12 Noise suppression improvement over single microphone case for interfering broadband noise at 90° for beamformed output (green), and voice isolation algorithm output (red).

suppressions existed. Note that where the voice signal is most prevalent, below 500Hz, the signal remains nearly unchanged while signals above 500Hz are suppression between 5dB and 30dB.

6.2.2.2 Interferer at 0°

For the last setup presented here, the interfering broadband noise source was moved to 0°, 6 feet in front of the manikin. This setup provided the challenges of a voice signal whose frequency range was masked by the interferer and a voice signal whose direction was the same as the interferer direction. The results from this final setup are shown in Figure 6.13. The amount of noise suppression for the 2 filtered cases is shown in Figure 6.14.

The performance of the GSC algorithm (red in both Figures 6.13 and 6.14), while slightly different than the first experiment setup, still suppresses the interfering broadband noise well and leaves the voice signal intact, except in the 300Hz range. The

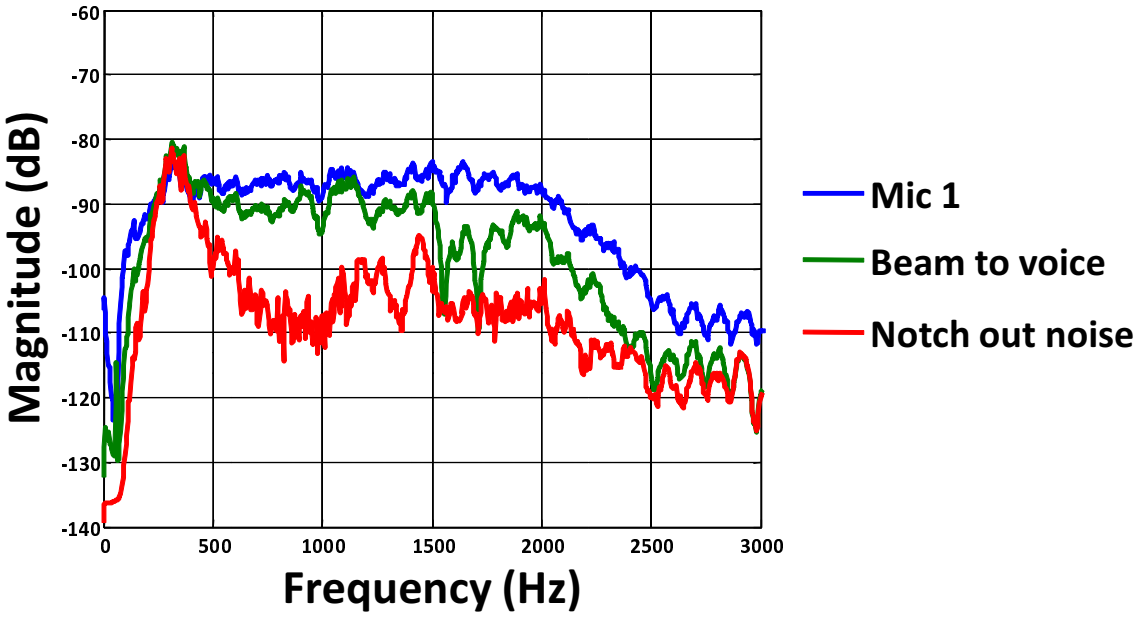


Figure 6.13 Second experiment setup results of voice isolation test with an interfering broadband signal for a single microphone signal (blue) and voice isolation algorithm output (red).

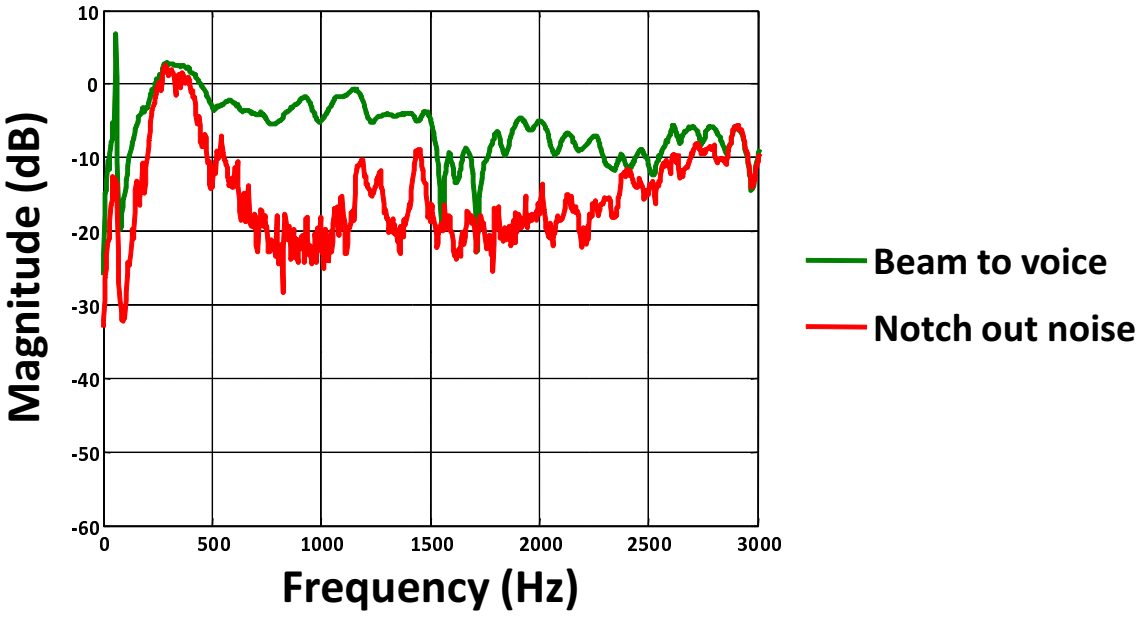


Figure 6.14 Noise suppression improvement over single microphone case for interfering broadband noise at 0° for beamformed output (green), and voice isolation algorithm output (red).

"beam-to-voice" beamformer does not provide the same level of performance when the interferer was changed from 90° to 0°. Suppression for the "beam-to-voice" beamformer was on the order of 5dB to 10dB for most of the frequency range with a maximum suppression of 22dB at 1700Hz. The improvement provided by the "block-voice" beamformer is marked by a suppression of 20dB for most of the operating frequency range with a maximum of 28dB at 800Hz. Paired with the results from section 6.2.1.2, this reinforces the ability of beamformers to "beam" in a direction with good results but "notch" in a direction and adaptively filter with even better results.

CHAPTER 7

Transparent Hearing Using Beampattern Matching

The culmination of array modeling, processing, and performance measurement is an optimization process for placing microphones in head mounted arrays. Since constructing a headphone array is not a trivial process, constructing many physical prototypes for testing all of the possible microphone array configurations is not feasible. This chapter details the computational model presented in Chapter 2 and the method by which it was used to test microphone configurations in a fraction of the time that real-world testing would require. The first section of this chapter provides a basis for how humans perceive and localize sound. A physical model of the human head and torso, called a KEMAR manikin, is then presented as the "target" for beampattern matching and as a platform for array testing. The chosen filtering strategy and several underlying assumptions are given as the main structure behind the optimization method. Having applied the optimization method, several prototype arrays are presented and their results compared to predictions from the computational model. Next, the layout of subjective tests using several filtering strategies and their results are presented with the occluded case as a control. Finally, the prototype array and the optimal filtering strategy presented are shown to provide statistically significant performance gains over the occluded sound localization case.

7.1 Biological Background

Many times advances in sensor arrays come about by seeking to imitate biological analogues such as echolocation and binaural hearing. The hearing in humans and animals is nothing short of amazing, especially when considering the ability of bats and dolphins to use sonar to accuracies rarely achieved by man-made systems. In the field of in-air acoustics, processing strategies follow the same methodology present in human hearing by exploiting time and amplitude differences experienced by sensors throughout the array. These differences that exist between the human acoustic sensors, the ears, are referred to as the interaural time difference (ITD) and interaural level difference (ILD). The relationship between sound from a single direction and the response measured at the human ears (ITD, ILD, and diffraction effects) is called a head related transfer function (HRTF). There are many other attributes to human hearing, perhaps so complicated that only as more is understood about brain activity and perception will more be understood about hearing as well.

7.1.1 Interaural Time Difference (ITD) and Interaural Level Difference (ILD)

The ITD and ILD used by humans to determine sound direction really describe two mechanisms used in different frequency ranges. For pure tones in the azimuth plane, the relationship between ITD and ILD and their roles in human acoustic localization is referred to as duplex theory, so named by Lord Rayleigh [56]. At low frequencies, where a quarter wavelength of sound is comparable to the diameter of the head, humans rely upon the ITD between the sounds heard at each ear [83]. At higher frequencies, where the incident wavelength is much smaller than the diameter of the head, humans use the ILD, or the intensity differences, between the sounds heard at each ear . Early physical studies of the human dependence on ITD and ILD have revealed that above approximately 1500 Hz the ITD is no longer employed and instead the ILD is used for sound localization [84][85].

As a simple example, consider a sound source directly to the right of the head. For a very low frequency sound (i.e. when the wavelength is very large compared to the head), the difference in sound level between the right and left ears is small, and the ITD is relied upon. A practical example of this is a subwoofer in a home theater system, where the location of the subwoofer does not affect the perception of the low frequency sound it transmits. Now if that same sound source is at a very high frequency (i.e. when the wavelength is very small compared to the head), the sound level at the right ear will be greater than at the left ear due to the head blocking, or shadowing, the sound, causing reliance upon the ILD. Using the home theater example, this is similar to the multiple small speakers placed around the room to provide spatial effects for higher frequency sounds such as bullets flying by, voices in a crowd, or traffic on a busy street.

While duplex theory provides an underlying basis for the functionality of human hearing, its simplicity lacks explanations for more complex hearing phenomena. For example, picture a plane that evenly divides the head into left and right halves, referred to as the median plane. Following duplex theory, human hearing should not be able to accurately localize sources placed anywhere on that plane. For a given sound source location to the side of the head (not directly collinear with the ears) there is a cone shaped locus of sound source locations that possess the same ITD and ILD, a phenomenon called the cone of confusion [86]. While some amount of localization ambiguity does exist, human listeners can correctly localize sources in both the median plane and for many "cones of confusion," suggesting that there are additional aspects to human hearing not accounted for by duplex theory. These particular aspects are thought to result from the acoustical scattering caused by physical structures of the human body.

7.1.2 Head Related Transfer Functions (HRTFs)

Head Related Transfer Functions describe the relationship between sound from any direction and the measured acoustic response at the ears. HRTFs depend on myriad unique properties of each individual's body from head size to nose shape, from the size of each ear to the angle that each ear makes with the head, from the width of the shoulders

to the length of the neck. While the ITD and ILD are embedded within, HRTFs also take into account the diffraction from, but not limited to, these aforementioned physical structures, known as spectral cues. As evidence that hearing may not be exclusively binaural (i.e. may not rely completely on ITD and ILD), early experimental results showed that human monaural localization was actually quite good [87][88]. Even experiments conducted later concluded "that monaural localization by persons totally deaf in one ear is commonplace," and that the structure of the pinna, or ear, performs a transformation on incident sound that the human brain uses when performing localization [89].

With the knowledge that HRTFs currently comprise the most complete and quantitative description of the way humans hear, stimulus synthesis techniques have been tested in experiments that compare localization of real sources to localization of virtual sources created through headphones [90]. The results of these experiments show that the spatial position of sound sources created over headphones are judged to have the same spatial position as real sources [91]. Another interesting finding of these experiments was that some humans may be predisposed to poor elevation discrimination, linked directly to the lack of elevation dependency of their measured HRTFs. In this way the front-back confusion and up-down confusion (both subsets of the "cone of confusion") predicted by the duplex theory exist to some extent, although the diffraction from physical features, while highly frequency dependent, does provide useful directional cues [92].

7.1.3 Other Human Hearing Methods

In addition to HRTFs, there are numerous strategies employed by humans to accurately localize sound in otherwise confusing situations or environments. One argument about the lack of front-back ambiguity found in humans in real world environments is the ability to use dynamic cues to correctly localize sounds [93]. These cues are produced by head movements or source movements with respect to the head, but in either case, result in greatly decreased incidence of front-back confusion [94]. In highly reverberant environments where localization would seem impossible due to multipath acoustics, humans can still determine the initial sound source direction, provided they were present at

the inception of the sound. This behavior, termed the precedence effect, shows that for sound sources temporally separated by less than 1ms, humans perceive both sources as a single "fused" source. For temporal separations greater than 1ms and less than 35ms, the leading sound source is perceived as the only sound source and is localized quite accurately. For temporal separations greater than 70ms, the sound events are perceived as completely separate and can be localized as such [95]. The phenomenon of undesirable echoes in telecommunications is an example of the precedence effect [96]. Litovsky et. al. provides a thorough review of the precedence effect and psychophysical research into human perception of temporally separated sound sources [97].

7.2 Measurement of HRTFs

While the HRTFs of individuals can be measured directly [90][91], a general physical model, called the KEMAR manikin, has been developed specifically for recreating the acoustic diffraction of the human body [98]. The KEMAR manikin consists of a torso section and a head section with removable pinnae inserts and openings at the entrances of the ear canals. While the main dimensions of the manikin body and head are unchangeable, the pinnae inserts allow measurement of the acoustical scattering caused by different pinna geometries. Figure 7.1 shows a close-up view of a left pinna insert.

Using the method described in Chapter 3.1, the HRTFs from a KEMAR manikin were used to obtain the performance "target" for a headphone microphone array [98]. The results of this measurement are shown in Figures 7.2(a) through 7.2(e) for the left ear for elevation angles of $\phi_m = -20°$, $0°$, $20°$, $40°$, and $60°$, respectively.

7.3 Optimization of Microphone Positions

7.3.1 Filtering Strategy

Selection of the best microphone locations for a headset array is directly tied to the beampattern matching strategy employed. Utilizing the frequency domain representation

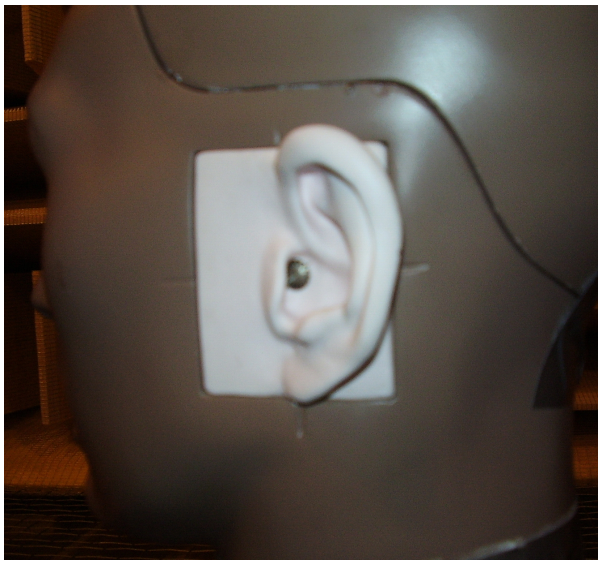
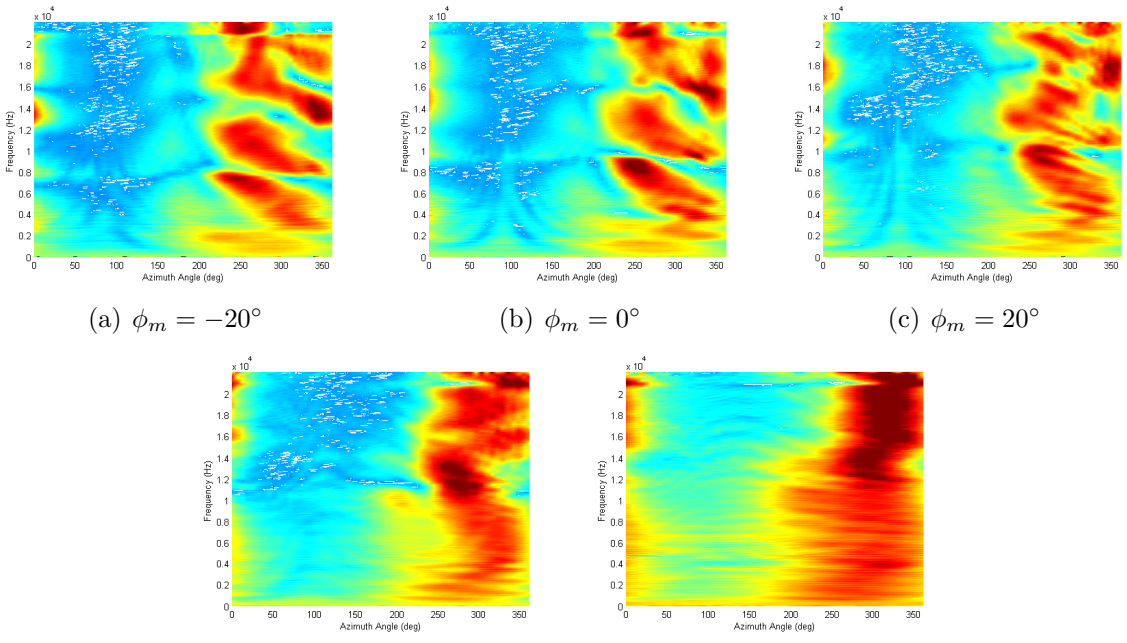


Figure 7.1 Left pinna insert in KEMAR manikin head.



(a) $\phi_m = -20°$ (b) $\phi_m = 0°$ (c) $\phi_m = 20°$

(d) $\phi_m = 40°$ (e) $\phi_m = 60°$

Figure 7.2 HRTFs measured from the left ear of a KEMAR manikin.

of the HRTFs produces magnitude and phase information at each frequency and source direction. For P frequencies and M source directions this produces $2PM$ values that must be matched for each ear to perfectly recreate the HRTFs. One specific difficulty involved in using headphone array signals to recreate HRTFs is the problem of acausality.

Consider a sound source to the right of the head located along a line drawn between the two ears. Sound from this direction will arrive at the right headphone microphones a fraction of a second earlier than it will at the right ear. That is to say that a filter applied to the microphone signals from the right headphone microphones, assuming that filter is applied instantaneously, will introduce a slight delay into the measured microphone signals to recreate the right ear HRTFs. Now consider a sound source located on the opposite side, to the left of the head. Sound from this direction, after diffracting around the head, will actually arrive at the right headphone microphones after it arrives at the right ear. That is to say that the filter will have to introduce a negative delay to accurately recreate the right ear HRTFs. More specifically, the measurement delay between the left and right headphone microphones is larger than the delay that exists between the left and right ears, requiring a negative delay in the applied filters to appropriately match the $2PM$ values for the HRTFs of each ear. Unfortunately the fact that headphones are generally always going to be wider than the human head, the problem of acausality will likely exist for nearly any array geometry.

Returning to the mechanisms of human hearing discussed in section 7.1 reveals that a perfect match of the magnitude and phase for every frequency and location may not be necessary. Although it does not encompass all parts of human hearing, duplex theory does well to describe the main spatial cues perceived by humans. The ITD used at low frequencies is captured by the phase of the HRTFS, while the ILD used at high frequencies is captured within the magnitude of the HRTFs. Matching only the phase at some frequencies and only the magnitude at other frequencies reduces the number of matches by a factor of 2 to PM.

7.3.2 Equivalent Source Headset Model

The equivalent source model from Figure 2.29 is used to reduce the number of candidate microphone mounting positions on the headphones. Some locations in the model are not ideal for microphone placement because of surface materials (i.e. the foam cups that seal on the sides of the head), underlying components (i.e. electronics), or size restrictions (i.e. small or fragile structures). Given these somewhat subjective criteria, the process of reducing the possible positions is a manual one. Figure 7.3 shows the total equivalent source model and the reduced set of candidate headphone points.

7.3.3 Simplifying Assumptions

Assume the simplification of HRTF matching where the magnitude is the only important concern. Given that the strategy is to minimize the error between the HRTF and the recreated HRTF in a least-squared sense, filter calculation seeks to match the magnitude and phase of the HRTF in a vector sense. If the magnitude of the HRTF is the only concern then the match or mismatch of phase can be ignored. Alternatively, if the HRTF is thought of as the "target" response, modifying the HRTF phase in a particular way may simplify or aid in filter calculation. This "adjusted target" response may allow better performance by creating filters that focus on differences that humans can perceive (i.e. magnitude) while neglecting compromises favoring differences that humans cannot perceive (i.e. phase).

The method used to modify the HRTF phase depends upon the filtering strategy used for transparent hearing. If the entire microphone array is used to recreate both left and right HRTFs then the system is coupled. However, if only the microphones on the left headphone are used to recreate the left HRTF and only the microphones on the right headphone are used to recreate the right HRTF then the system is uncoupled. The difficulty with assuming a coupled system is the introduction of acausality into filters (i.e. sounds cannot be presented at the ears until they have been received by microphones on both headphones, see section 7.3.1). Conversely, assuming an uncoupled system removes

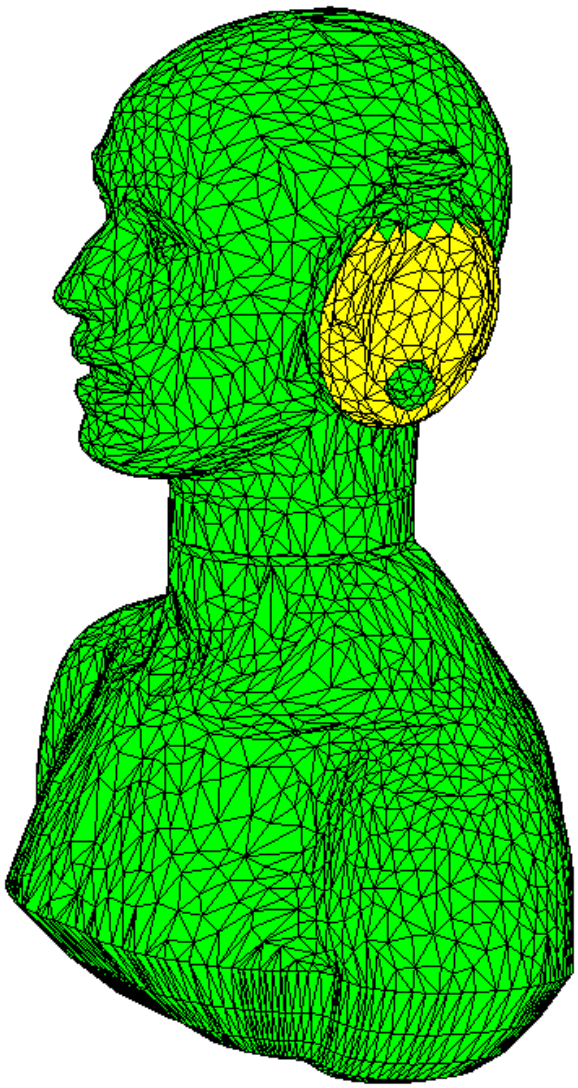


Figure 7.3 Equivalent source model of KEMAR and headset (green) and candidate microphone locations (yellow).

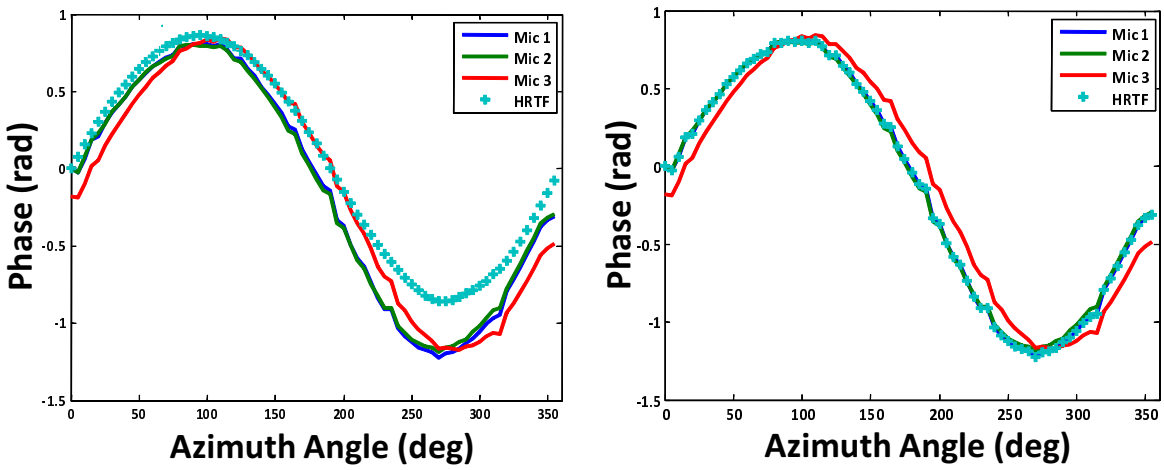
many situations that require acausal filtering. For the purposes of filter calculation presented here, the filtering for the left and right ears is assumed to be completely uncoupled.

With the aforementioned assumptions in mind, the modification of the HRTF phase is performed based on the phase of microphone signals. Specifically, the phase of the HRTF for one side will be adjusted to match the phase of a single microphone on the corresponding headphone. This approach acknowledges the phase differences that exist between the microphones based on different sound source locations, but adjusts the HRTF phase of that side to allow a match of the magnitude. At a given frequency, the phase of the HRTF for the i_{th} angle takes the phase of first microphone at the i_{th} angle while maintaining its magnitude.

$$D_{HRTF_{adjust},i} = |D_{HRTF,i}|e^{j\angle D_1,i} \tag{7.1}$$

The results shown in the Figures 7.4 through 7.6 come from experimental results with the headset array presented later in Figure 7.13. As an example, Figure 7.4(a) shows the original phase measured for 3 headphone array microphones and the left ear HRTF. Notice that the HRTF phase has less amplitude than any of the individual microphone phases. This means, using the terms phase and delay interchangeably, that there is a larger delay inherent in the microphone measurements than is present in the HRTF. Unlike adjusting the magnitude to match, the shorter delay present in the HRTF measurements can never be compensated for by the delays or any sum of the delays present in the microphone measurements. The compensation literally would only work if the microphones could detect incoming sound before it arrives. Figure 7.4(b) shows the same phase for the 3 headphone array microphones, but with the phase of the left ear HRTF adjusted to match the first microphone.

The fact that human perception may not be able to discern the difference between the original HRTF and the adjusted HRTF is what allows such a phase adjustment to be performed. Extending the example from Figure 7.4, the error between the recreated HRTF and the original HRTF is shown in Figure 7.5(a) while the error between the

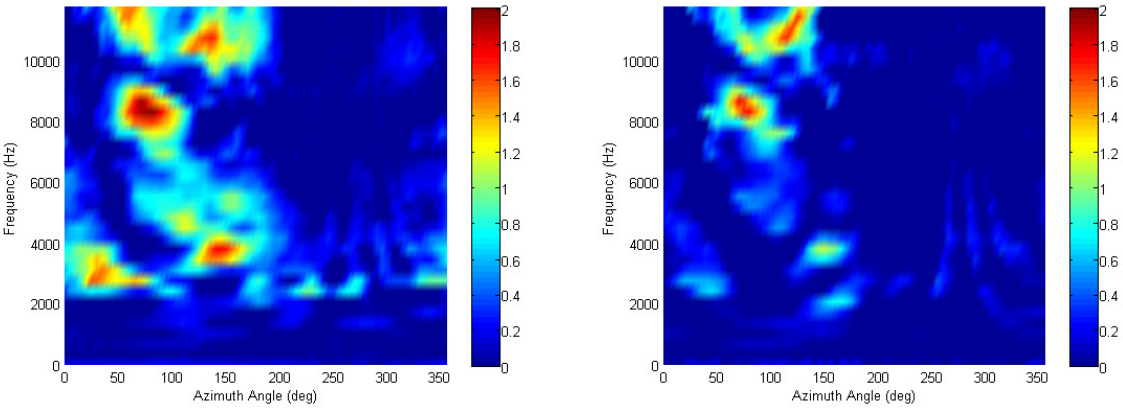


(a) Original phase

(b) Phase adjusted

Figure 7.4 Adjustment of HRTF phase before calculation of filters.

recalculated recreated HRTF and the adjusted HRTF is shown in Figure 7.5(b). Note that the illustrations used here consider only the 0° elevation for the sake of simplicity.



(a) $|D_{HRTF}| - |D_{recreated_1}|$

(b) $|D_{HRTF_{adjust}}| - |D_{recreated_2}|$

Figure 7.5 Magnitude error between recreated HRTF's and the original and phase adjusted left ear HRTF's at 0° elevation.

7.3.4 Selection of Microphone Positions

Using the results from the equivalent source model shown in Figure 7.3 in conjunction with the filtering method described in sections 7.3.1 and 7.3.3 allows the selection of

optimal microphone locations. There were 248 possible microphone positions on the left headphone in the computational model suitable for creating an array. Assuming 3 microphones per headphone and the filtering strategy previously mentioned, there are several million possibly microphone configurations. Each possible array geometry was evaluated based upon the positions of the microphones and the order in which the microphones were placed. The order mattered because the HRTF phase is adjusted based on the first microphone in the array while normal filtering methods are applied to the second and third microphones. The resulting number of possible microphone array geometries for the left headphone is,

$$\frac{(248)(248-1)(248-2)}{2} = 7,534,488 \tag{7.2}$$

Through an exhaustive search of all possible array geometries the optimal geometry for the given filtering strategy was determined. Note that the optimal geometry is highly dependent upon the cost function used to evaluate each array's performance. For this search the cost function was the percent difference squared between the recreated HRTF and the adjusted HRTF, summed across all directions at each frequency. The results of this calculation are shown for all frequencies used in Figure 7.6.

$$J_i = \frac{|D_{HRTF_{adjust}}(f_i) - D_{recreated}(f_i)|^2}{|D_{HRTF_{adjust}}(f_i)|^2} \tag{7.3}$$

A single measure of performance was obtained by summing all cost function results across all frequencies,

$$J_{total} = \sum_{i=1}^{I} J_i \tag{7.4}$$

The results of the exhaustive search, sorted in ascending order, are shown in Figure 7.7. The optimization was performed over all azimuth and elevation angles available (308 positions total). The minimum of the cost function at each frequency is bounded by 0 while the maximum is bounded by 1. With a sample size of 35 frequencies ($I = 35$) the maximum possible error is 35 for each microphone array geometry. The optimal

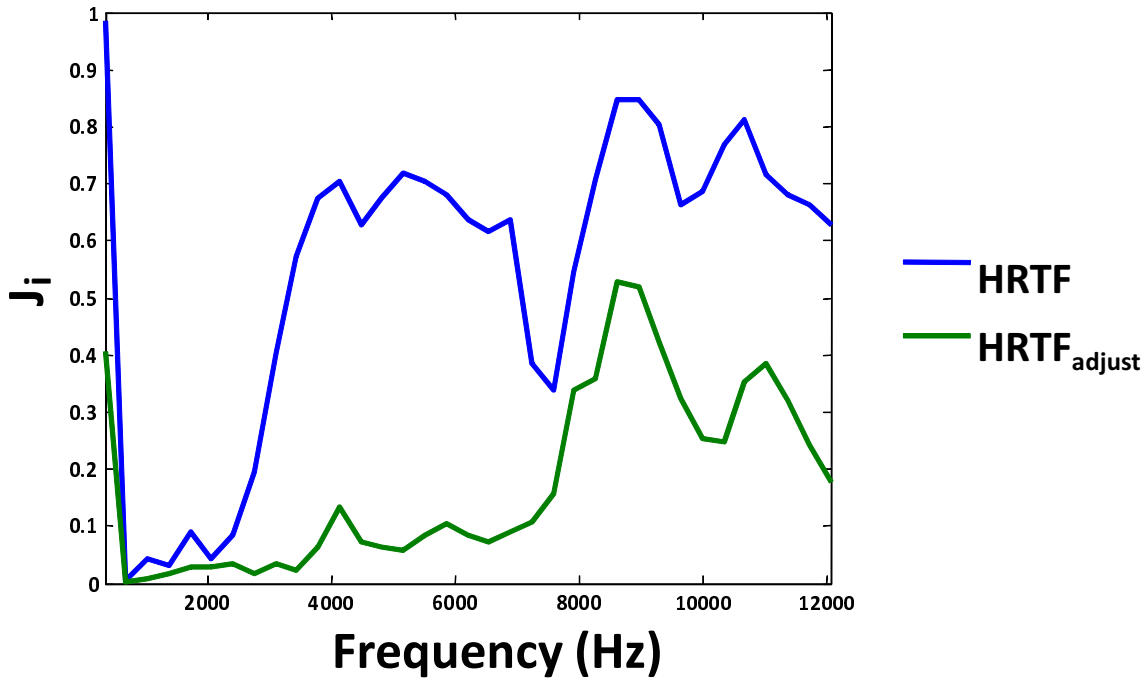


Figure 7.6 Error between recreated HRTF's and the original and phase adjusted left ear HRTF's at 0° elevation, summed across all directions for each frequency.

microphone array geometry for matching the left ear HRTF is not shown due to a non-disclosure agreement with Sennheiser. An alternative to an exhaustive search is to first select the microphone that best recreates the HRTF response then select the second and then third microphones to further refine the HRTF match. For the 248 microphone positions used in this application the number of iterations is reduced from over 7 million to 741. In both cases the same array geometry yields the optimal result.

The optimal microphone array geometry attempts to mimimize the cost function across all azimuth and elevation angles. Since human hearing is much more accurate at discerning azimuth differences than it is with elevation differences, the adjusted array geometry in Figure 7.8 was selected that would afford better discrimination between azimuth differences [99]. From the results of the exhaustive search, this new array geometry performs within the top 10,000, or the top 0.1%, of all possible array configurations as shown in red in Figure 7.7.

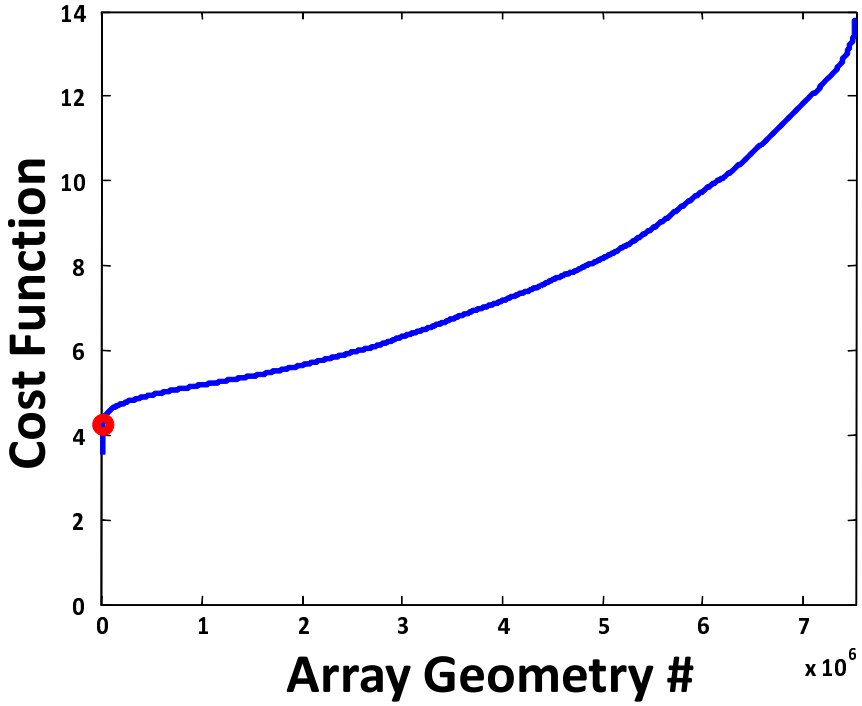


Figure 7.7 Cost function for all possible array geometries with compromise geometry shown (red). Geometries are sorted in ascending order based on error.

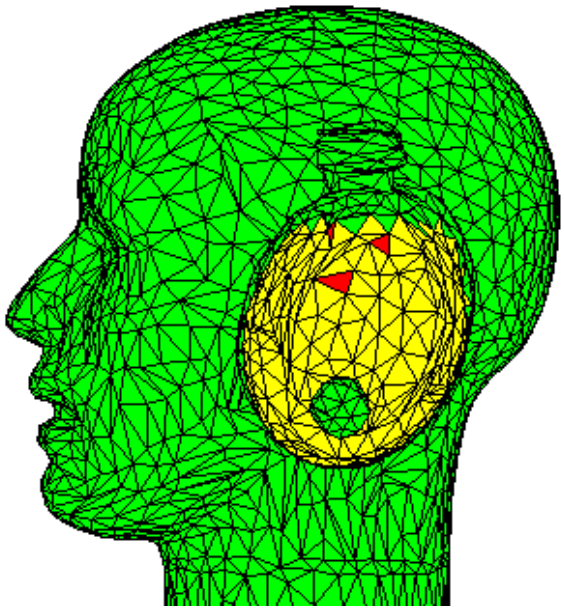


Figure 7.8 Alternative 3 microphone array geometry (red) for matching the left ear HRTF.

7.4 Performance of Arrays

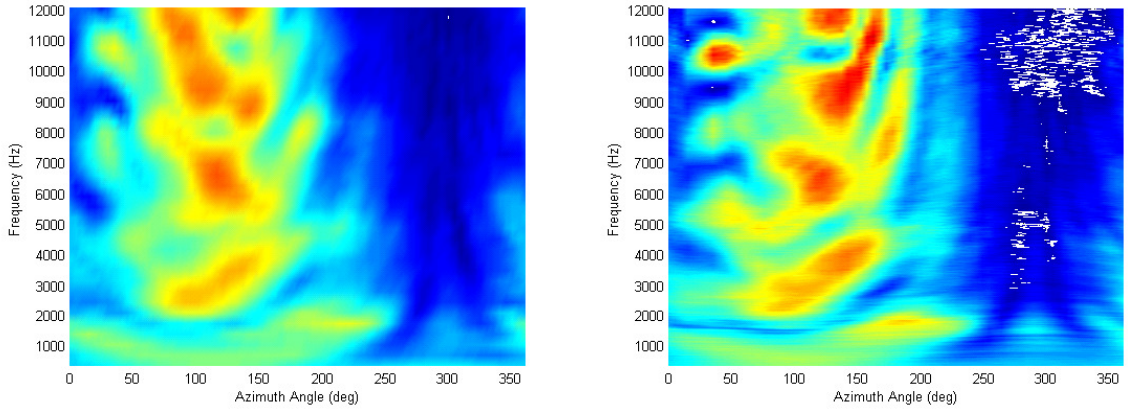
Optimization methods and strategies were validated by testing headset array prototypes. Quantitative test results indicate how closely computational models match real-world systems while qualitative test results provide a measure of how well filtering strategies achieve their goals. While analyzing quantitative results is straightforward, qualitative results can come in categories of feeling, comfort, and overall impression. This is to say that a real-world system may closely match theory from a quantitative view but be completely unusable due to its interaction (or lack thereof) with human perception. The following sections examine the results of testing array prototypes objectively and subjectively.

Collaboration with Sennheiser GmbH and its North American research group in Palo Alto, California fostered creation of headset arrays on Sennheiser HMEC-250 headsets. The first headset array was designed to match the optimal array geometry as closely as possible. The second headset array used the geometry shown in Figure 7.8 with raised microphones. Availability of sensors at the time of construction mandated that these 2 prototypes be built with electret microphones.

7.4.1 Optimal Geometry

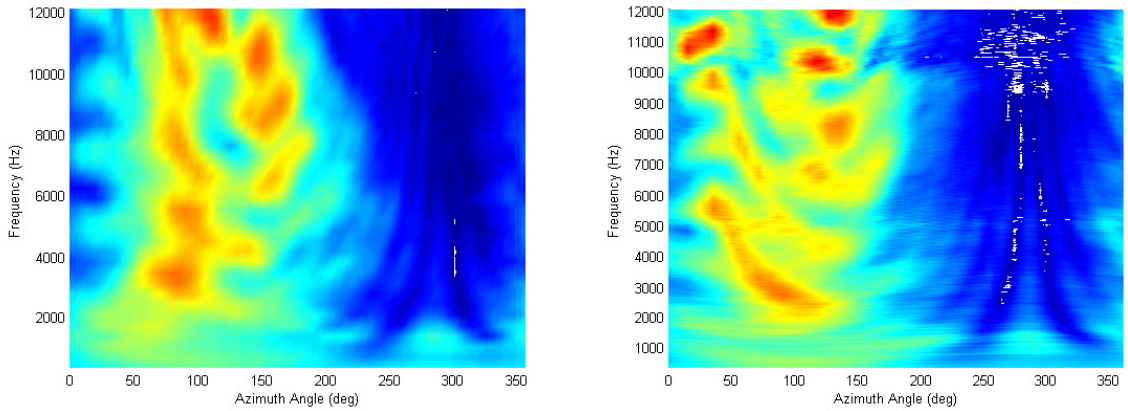
The experimental validation of the computational model results began with the construction of an array following the optimal array geometry. This array was built without regard for physical limitations of the headset, i.e. all other constraints were compromised in favor of obtaining the optimal array geometry. However, the underlying headphone structures (circuit boards, switches, speakers, etc.) ultimately caused a couple of the microphones to protrude approximately 2mm from the headphone surface. The computational model was adjusted to reflect the change in these microphone positions and these "refined" model results are presented in this section. The constructed array is not shown due to a non-disclosure agreement with Sennheiser. Comparisons of the array

responses from the computational model and experimental measurements are shown for an elevation of 0° in Figures 7.9, 7.10, and 7.11.



(a) Model (b) Experiment

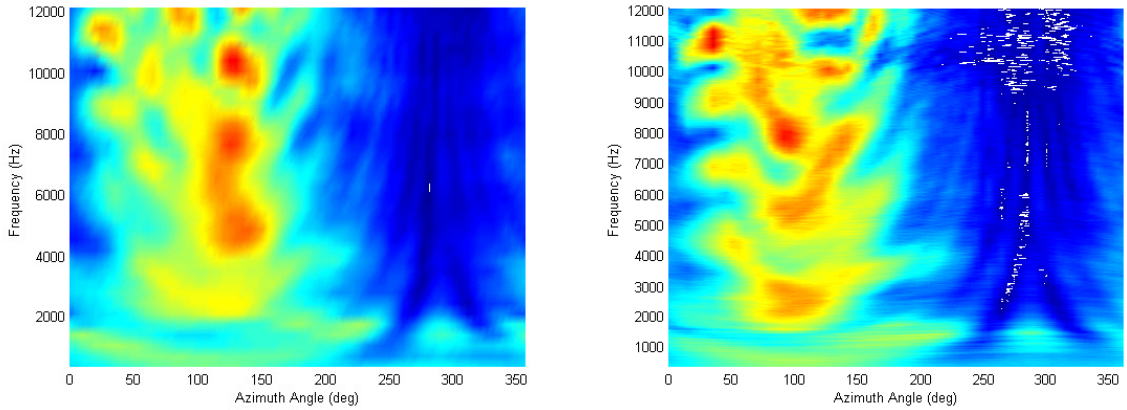
Figure 7.9 Comparison of computational model and experimental results for microphone 1 on the left headphone at 0° elevation for the optimal array geometry.



(a) Model (b) Experiment

Figure 7.10 Comparison of computational model and experimental results for microphone 2 on the left headphone at 0° elevation for the optimal array geometry.

Using Figure 7.6 as a template, the expected performance from computational models and experimental measurements is compared. The original computational model, experimental measurements, and refined computational model were all used to examine the predicted HRTF match error summed over all directions for each frequency, as shown



(a) Model (b) Experiment

Figure 7.11 Comparison of computational model and experimental results for microphone 3 on the left headphone at 0° elevation for the optimal array geometry.

in Figure 7.12. The comparison of these 3 errors reveals that the refined computational model provides a better match than the original computational model. Overall, the comparison between experimental results and both computational models shows excellent agreement in potential array performance.

7.4.2 Raised Microphones

Due to the underlying structures and size constraints of the microphone hardware, the microphones were mounted on small structures that offset them from the headset surface, as shown in Figure 7.13. From a prototype construction standpoint this method allows a relatively short build time in preparation for testing. However, since the computational model did not contain these microphone mounting structures, the real-world and computational results were measured from different diffracting bodies. Nonetheless, the comparison between computational and prototype results can be made, as shown in Figures 7.14, 7.15, and 7.16.

The compromise made to calculate the raised microphone array response was to use the computational model presented in section 7.3.2 to find the pressure field at the 3-space points where the microphones were located. The model was not modified to include the

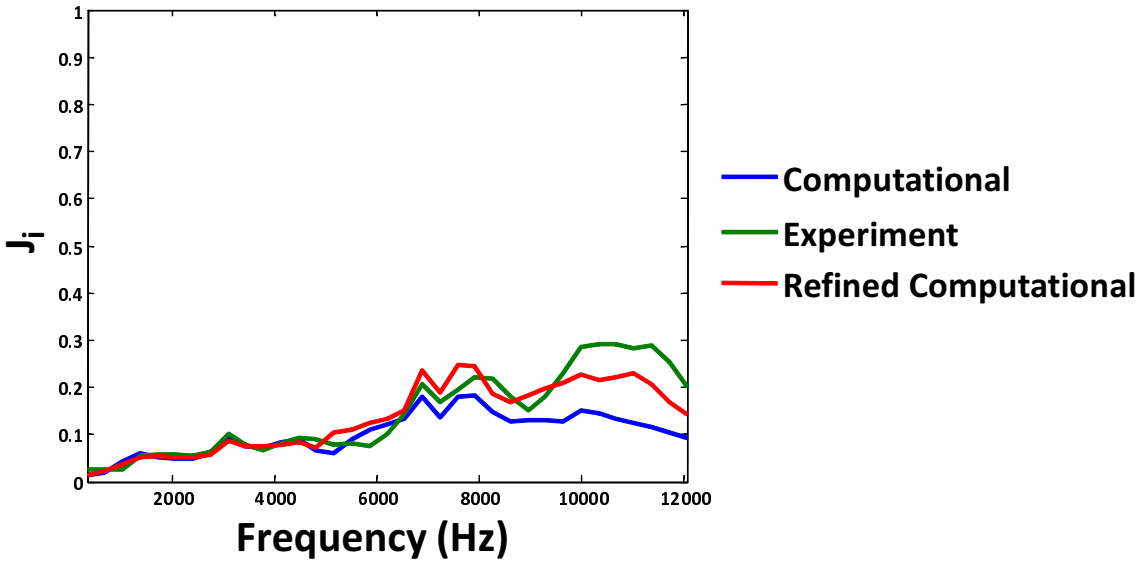
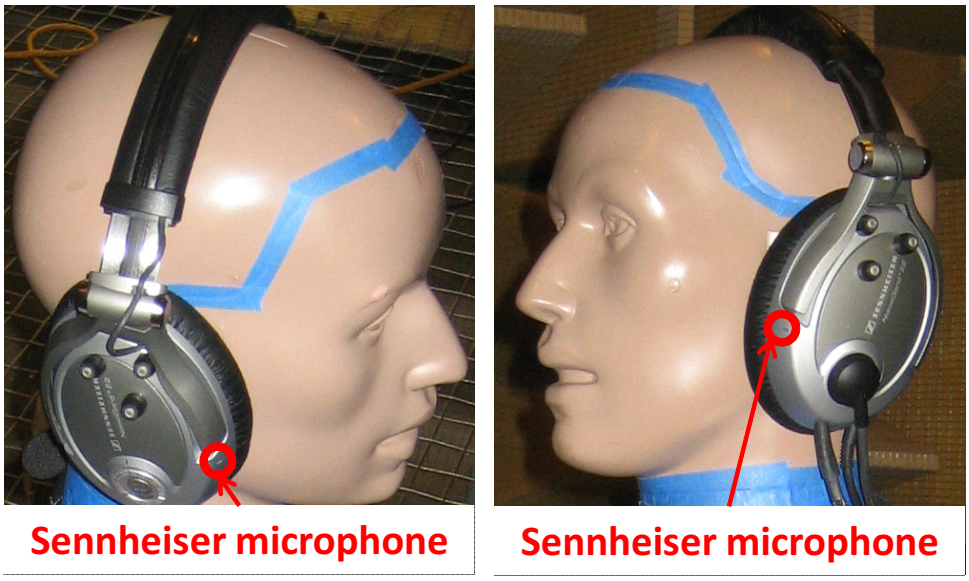
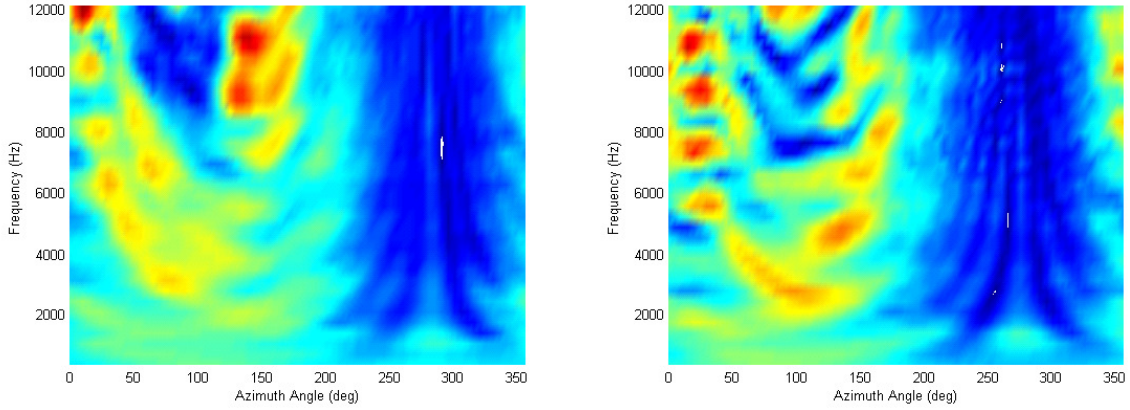


Figure 7.12 Error between recreated HRTF's and phase adjusted left ear HRTF's using the original computational model (flush-mounted microphones), measurements from the optimal geometry headset, and the refined computational model (positions reflect microphone locations on prototype), summed across all directions for each frequency.



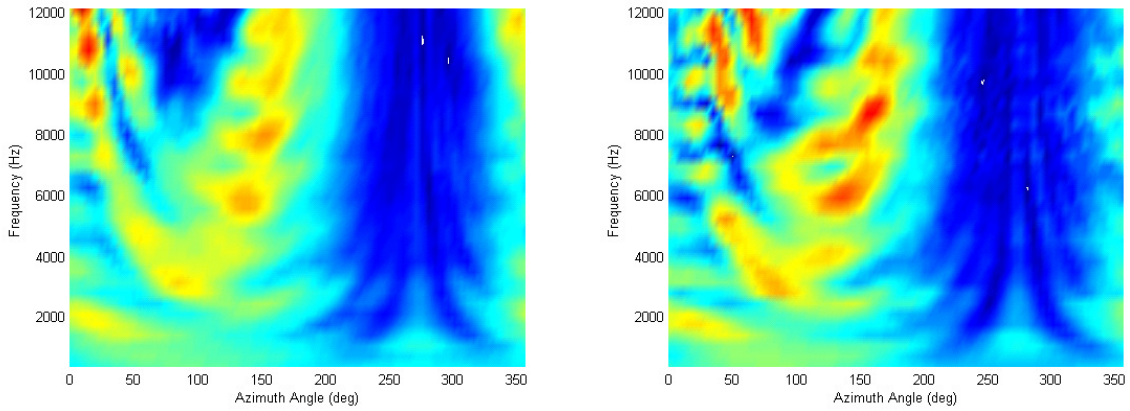
(a) Right headphone (b) Left headphone

Figure 7.13 KEMAR manikin wearing electret microphone headset array using raised microphone mounts.



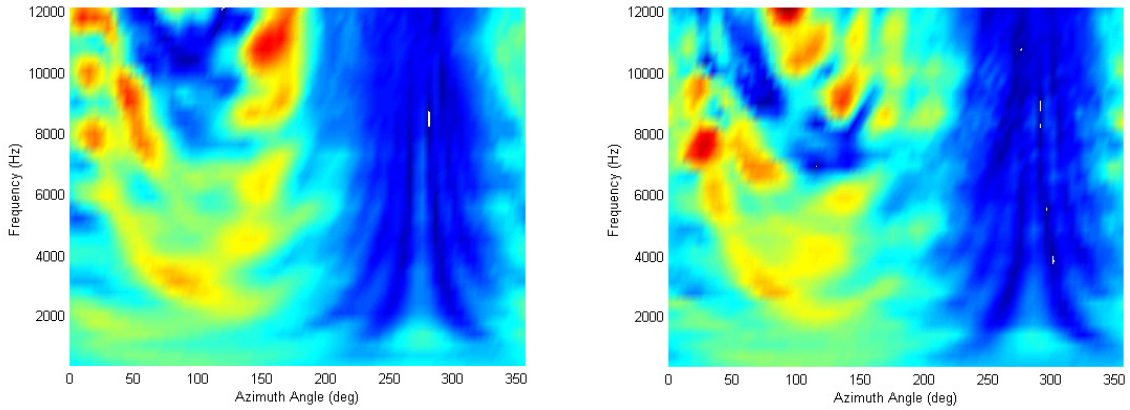
(a) Model (b) Experiment

Figure 7.14 Comparison of computational model and experimental results for microphone 1 on the left headphone at 0° elevation.



(a) Model (b) Experiment

Figure 7.15 Comparison of computational model and experimental results for microphone 2 on the left headphone at 0° elevation.



(a) Model

(b) Experiment

Figure 7.16 Comparison of computational model and experimental results for microphone 3 on the left headphone at 0° elevation.

microphone mounting structures, so the diffraction from those structures was neglected. However, the displacement of the microphones from the reflecting surface of the headphones causes a notch in the measured response at a predictable frequency corresponding to where one quarter wavelength of sound is equal to the length of the structures. This notch arises when sound waves reflecting off the headset surface destructively interfere with incoming sound waves at the microphone position. The expectation is that the notch occurs when a microphone faces the sound direction (around 90° for microphones on the left headphone and 270° for the right headphone) and the wavelength meets the previous criteria (one quarter wavelength corresponds to 9mm, the height of the microphone mounts, or approximately 9200 Hz). The results show that the large features of the responses are captured, but that the finer detail is not matched as well. For example, the shadowing present between azimuth angles of 200° and 360° are duplicated quite well, but the response peaks are not as intense or slightly shifted in angle and frequency compared to what the computational model predicted.

The expected performance from the "raised" microphone computational model and experimental measurements is shown in Figure 7.17. As in the case of the optimal geom-

etry, the potential error between both computational and experiment results provides a good match.

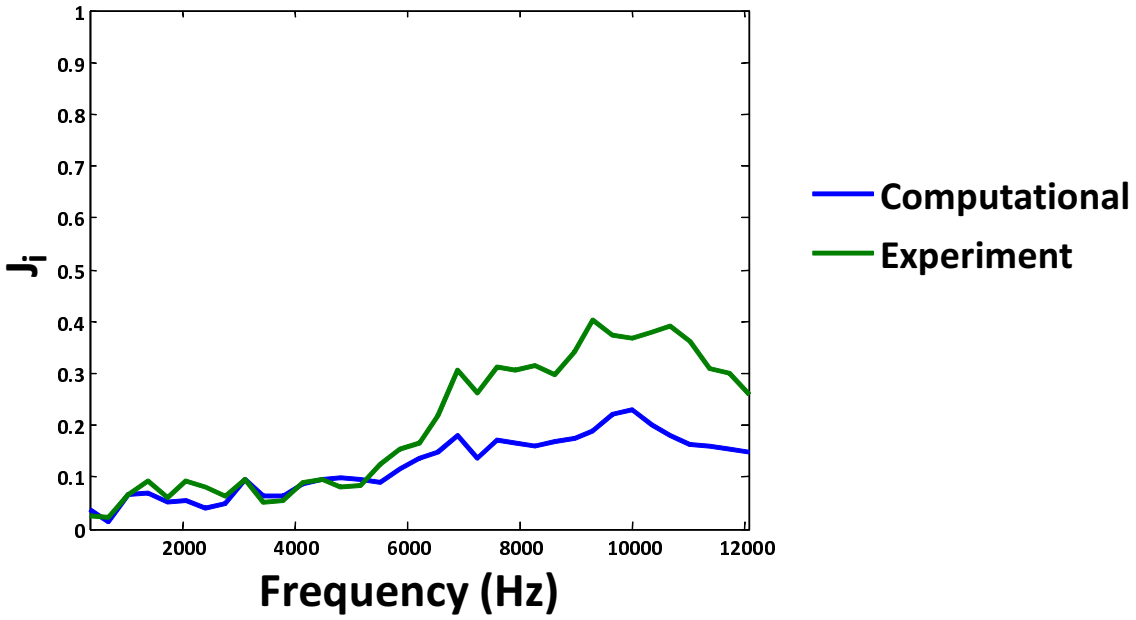
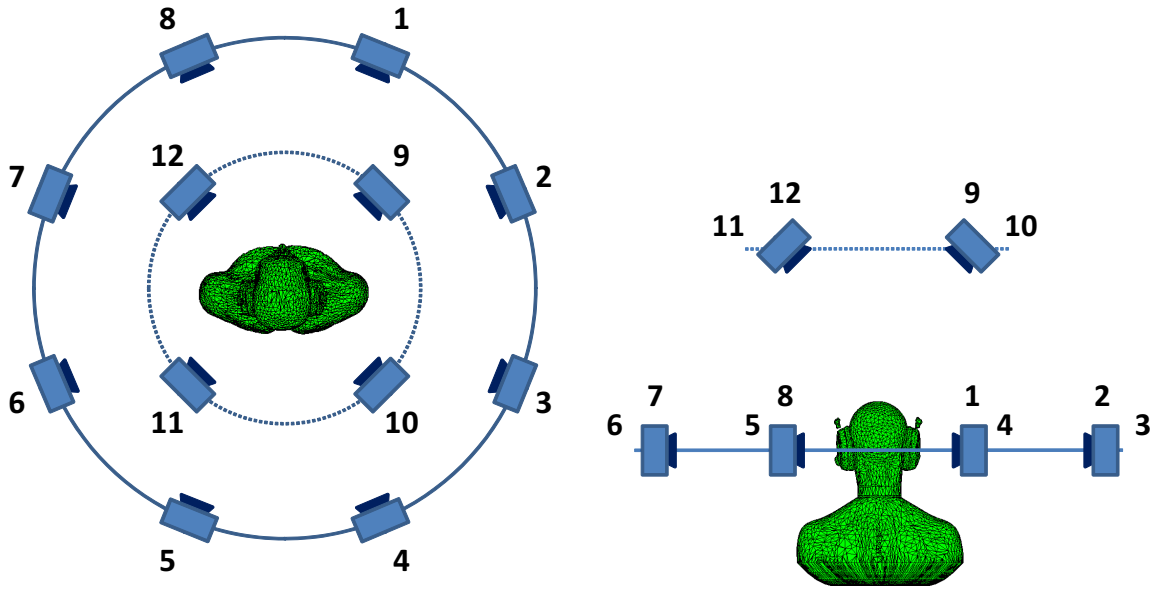


Figure 7.17 Error between recreated HRTF's and phase adjusted left ear HRTF's using the "raised" microphone computational model and experimental measurements, summed across all directions for each frequency.

7.5 Transparent Hearing Experimental Layout

Experiments were conducted to test the filtering strategies and the feasibility of the transparent hearing application. The room used for testing was equipped with 12 speakers split up into 2 rings; 8 speakers in the 0° elevation ring and 4 speakers in the 45° elevation ring. The speaker setup, while not drawn to scale, is shown from an overhead view and a rear view in Figure 7.18. A subject participating in an experiment is shown in the actual testing environment in Figure 7.19 with speakers 1, 2, and 9 highlighted in white. One benefit of this speaker/subject positioning is that speakers 1, 4, 9, and 10 lay on a "cone of confusion" (section 7.1.1) for the right side (as do speakers 2 and 3) while speakers 5, 8, 11, and 12 do the same for the left side (as do speakers 6 and 7). Figure 7.20 illustrates

a “cone of confusion” for the left side where any sources positioned on the solid red line will have identical ILD’s and ITD’s.



(a) Overhead view (b) Rear view

Figure 7.18 Overhead and rear view of speaker setup for transparent hearing experiments.

7.6 Subjective Tests

Since HRTF’s and the headset array manifold matrices are based on all the physical attributes of a person’s head, neck, torso, ears, etc., they are likely to be different for each person. Establishing the need for individualized HRTF’s and headset array manifold matrices and the current level of transparent hearing technology were the main goals of human subject testing. For each subject undergoing psychoacoustic testing, a set of HRTF’s and headset array manifold matrices were measured using 12 sound directions (see Figure 7.18). This provided a set of “individualized” measurements that were then averaged into a set of “generic” measurements, each implemented in different test cases. For each human subject there were 5 test cases:

- Occluded (wearing headset with no active transparent hearing system)

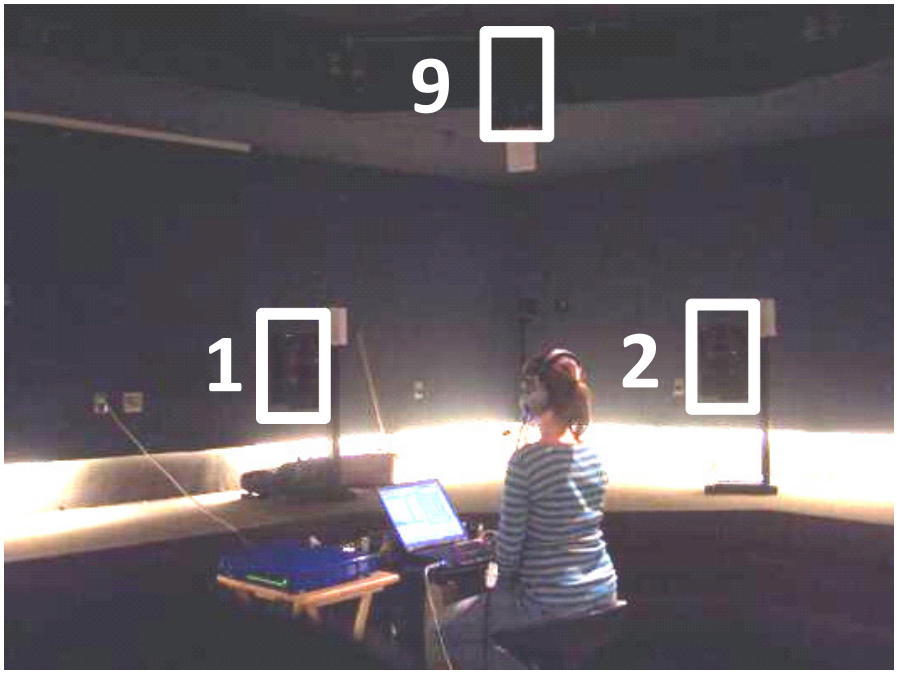


Figure 7.19 A human subject participating in transparent hearing tests with speakers labeled (numbered white boxes).

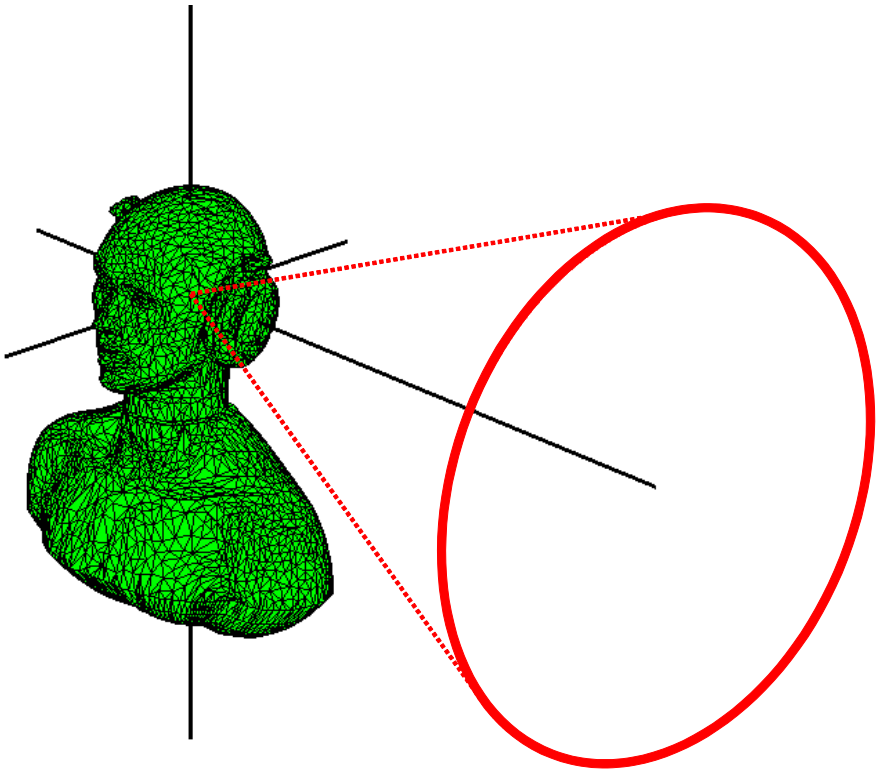


Figure 7.20 A sample cone of confusion (solid red) for the left side of the head.

- Talk-through (utilizing Sennheiser factory equipped system - a microphone on each headphone feeds sound through (see Figure 7.13))

- 3-microphone individualized HRTF's and headset array manifold matrices (3 microphones per headphone with measured HRTF's and headset array manifold matrices for each subject and the filtering strategies from section 7.3.1)

- 3-microphone generic HRTF's and headset array manifold matrices (same as previous strategy but with a set of HRTF's whose frequency responses were in the middle of the subject group and headset array manifold matrices applied to all subjects)

- 1-microphone generic HRTF's and headset array manifold matrices (assessing filtering strategy performance with only 1 microphone per headphone)

For each case a 5/8 second long modulated white noise burst, shown in Figure 7.21, was played 36 times (3 times per speaker) in a random order. By modulating white noise the frequency content of the stimulus remained flat while multiple impulses were presented. The subject was asked to localize each burst through an automated computer program that recorded all results.

During the testing, the active noise cancellation system (called *NoiseGuard*) was turned on and the listener was specifically requested to rotate his or her head to aid in sound localization. Therefore, the dynamic cues previously mentioned (section 7.1.3) were not isolated and tested during experiments. In a real world environment it is unlikely that a transparent hearing user would be unable to move his or her head and the experiments were meant to imitate these real world circumstances.

7.7 Test Results

The results of the subjective tests, shown in Table 7.1, are described in terms of the number of incorrect source locations chosen by each subject. Shown again in Figure 7.22,

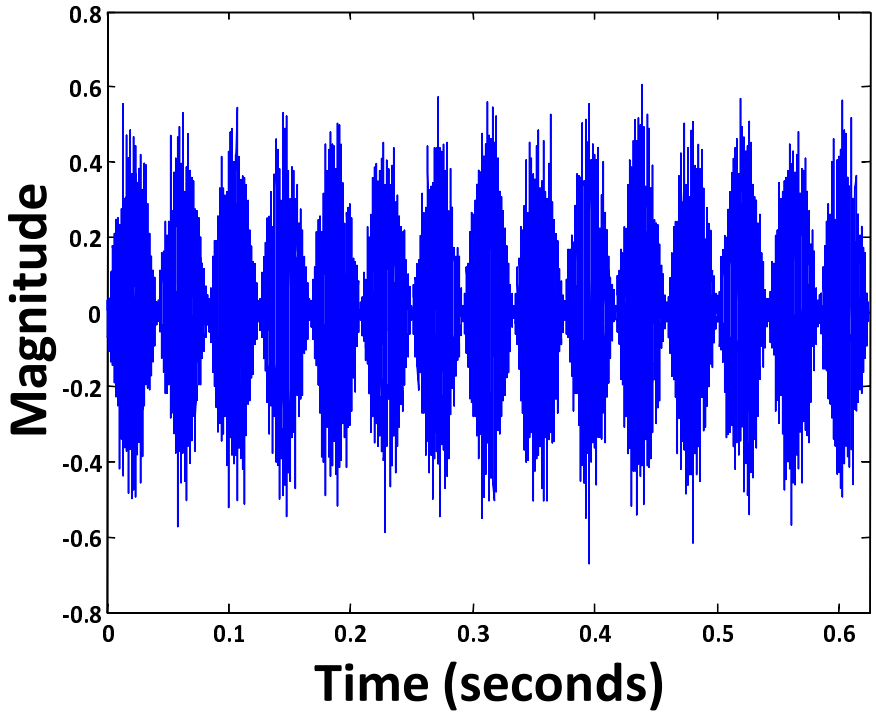


Figure 7.21 Modulated white noise burst stimulus used for transparent hearing testing.

the results indicate that the individualized 3-microphone transparent hearing case may have mitigated some localization errors introduced by occluding the ears.

Another way to view the data is in terms of the number of "cone of confusion" errors. These results, shown in Table 7.2, examine when incorrect sound localization occurred within the left side subset of speakers 5, 8, 11, and 12 (or 6 and 7), or within the right side subset of speakers 1, 4, 9, and 10 (or 2 and 3). As in the overall results, the "cone of confusion" results indicate that the individualized 3-microphone transparent hearing case may have improved the directional discernment of sources.

While some difference between the cases exists, whether or not the difference is statistically significant remains a question. Using the occluded results as a baseline for all other results, the statistical significance of the performance improvement is calculated. The mean difference and 95% confidence interval bars are plotted for the overall results in Figure 7.23 and the "cone of confusion" results in Figure 7.24. Since the statistics are based on the differences between each case and the occluded case, the statistical test

Table 7.1 Number of incorrect source localization responses from each subject.

Subject	Occluded	Talk-through	Individual 3	Generic 3	Generic 1
1	13	12	7	4	16
2	11	5	4	7	4
3	20	11	11	11	15
4	15	7	1	9	9
5	13	13	14	11	18
6	17	16	16	14	21
7	17	20	13	19	16
8	7	8	3	5	9
9	9	10	4	7	16

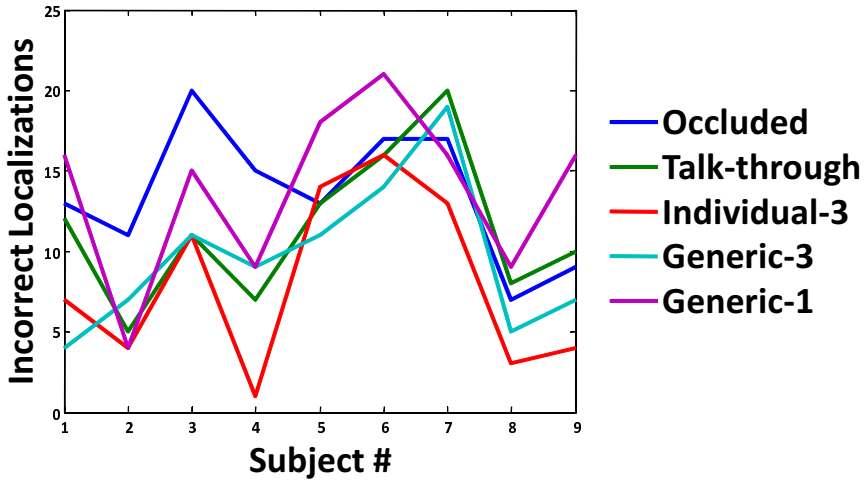


Figure 7.22 Results of subjective experiments for 5 cases and for each subject.

Table 7.2 Number of incorrect "cone of confusion" source localization responses from each subject.

Subject	Occluded	Talk-through	Individual 3	Generic 3	Generic 1
1	10	5	5	2	12
2	10	4	3	6	4
3	17	5	4	5	9
4	11	4	0	6	5
5	9	6	10	6	16
6	13	9	6	5	14
7	12	13	5	16	11
8	7	5	3	4	9
9	8	6	3	4	9

is whether the difference in the means is 0. When the confidence interval bars do not cross 0 for a particular case, the null hypothesis, that the difference in the means is 0, is rejected (i.e. the improvement or digression for that case is statistically significant).

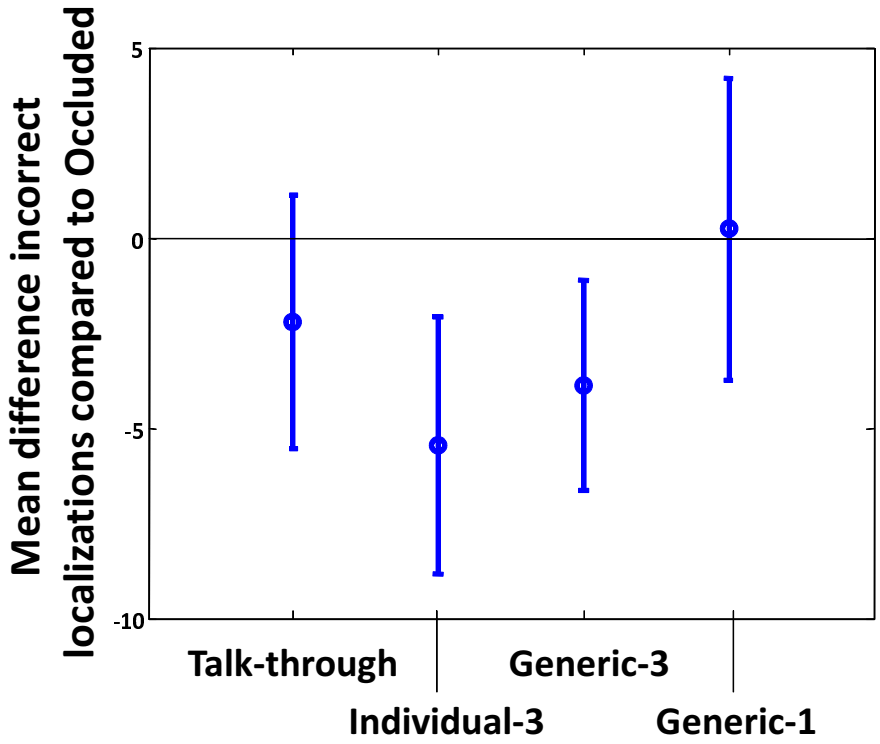


Figure 7.23 Mean difference and confidence intervals for determining improvement presented by each method for overall results.

Figure 7.23 shows that only the improvements offered by the individual and generic 3-microphone array strategies are statistically significant. More specifically, the p-values for these 2 methods were respectively 0.0029 and 0.0058, both less than the 0.05 level indicated by the 95% confidence interval. For the subset of the "cone of confusion" performance, Figure 7.24 shows that, additionally, the talk-through strategy performance is statistically significant. For the individual-3, the generic-3, and the talk-through strategies, the p-values are 0.0023, 0.006, and 0.0035, respectively.

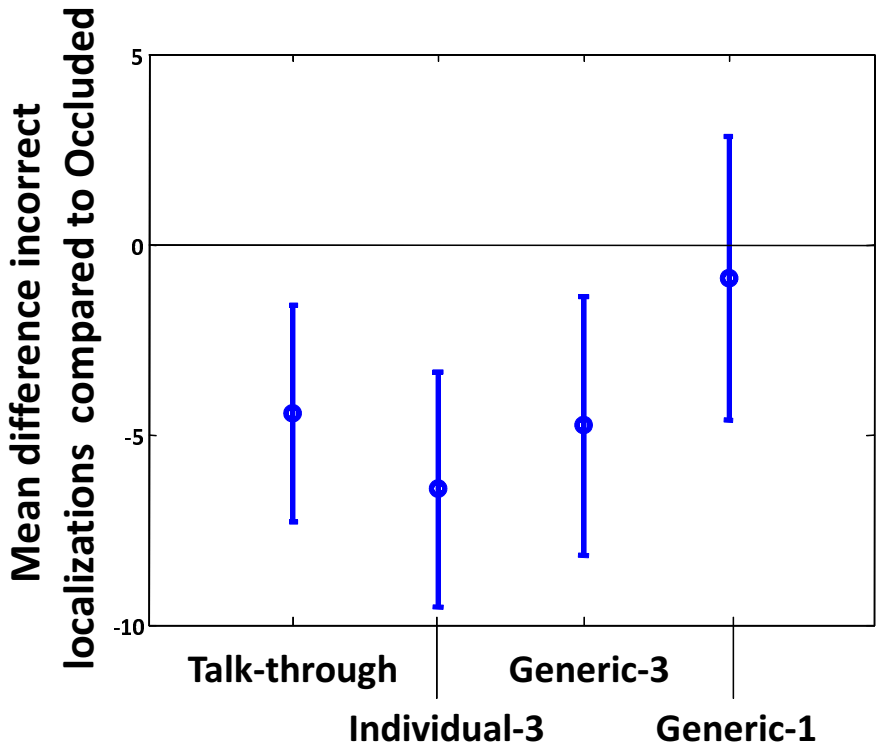


Figure 7.24 Mean difference and confidence intervals for determining improvement presented by each method for "cone of confusion" results.

CHAPTER 8

Summary and Future Work

The main objective of the work presented in this thesis was to study the use of head mounted microphone arrays used for a variety of applications. More specifically, it was important to develop methods for characterizing their behavior, evaluating their performance and optimizing their design. This thesis began with the characterization of microphone arrays using a manifold matrix approach and a number of models of increasing complexity were investigated using both theoretical and numerical models and using experiments. The final model used was an equivalent source diffraction model using a 8824 face scanned KEMAR manikin wearing a HMEC-250 Sennheiser headset. This model was shown to be accurate up to 12KHz by comparing model results with experimental measurements of a microphone array mounted on the HMEC-250 headphones. The model was used with an array geometry optimization strategy based upon the limitations of human hearing and the application of transparent hearing. Both an optimal and a modified optimal geometry were used to construct headset arrays and those arrays were tested experimentally to further validate model results. The modified optimal geometry was used in subjective testing to determine the performance improvement over currently available transparent hearing techniques. Finally, subjective testing results showed that the improvement in performance over other tested methods was statistically significant. The following sections summarize the contributions of the work and presents ideas for further research.

8.1 Performance of dissimilar arrays

The performance metrics developed in Chapter 5 provided ways of comparing the performance of arrays with different geometries. Specifically the Directivity Index (DI) and White Noise Gain (WNG) were used as ways of evaluating an array's ability to find sources in a diffuse noise field and to evaluate the array's resistance to sensor noise. It was shown that a conditioning factor alpha, used in the calculation of the optimal array filters, could be varied to tradeoff DI against WNG. The singular values of the sensor's cross spectral density matrix were observed to have a large impact on this tradeoff.

It was shown that the performance of free field and diffracting arrays of identical geometry is not the same. The increased performance can be viewed in two ways: (i) at a given frequency the diffracting array has a better DI and sensitivity to noise (WNG) than an equivalent non-diffracting array or (ii) a non-diffracting array will have to be physically larger to achieve equivalent performance. For example at $ka = 0.2$ the diffracting cylindrical array achieved 5dB improvement in WNG over a non-diffracting array with similar DI or the non-diffracting array would have to have a diameter 55% larger in order to achieve the same performance. This improved performance due to the presence of a diffractor was also demonstrated for spherical arrays both numerically and experimentally. This increase in performance was shown to be due to the increased magnitudes of the singular values of the CSD matrix. It was shown that at low frequencies the diffracting array always had larger singular values as compared to a non-diffracting free-field array. "Knees" in the WNG-DI tradeoff curves were shown to be due to the inclusion (or exclusion) of groups of singular values.

The methodology developed here allows the performance of arrays of different geometries to be compared. For example, the WNG-DI tradeoff curves allow the performance over a wide range of designs to be evaluated quickly and help find appropriate WNG-DI tradeoffs. Specifically, the increased performance of diffracting arrays over non-diffracting arrays implies that head mounted arrays will actually outperform their free-field equivalents especially at low frequencies. This is important for the application of long range

detections where low frequency signals dominate (high frequencies attenuate faster long distances).

The ability to compare the performance of arrays regardless of geometry will lead to a better understanding of array design and promote development of more design tools to optimize arrays before construction. The role that diffraction plays in improving array performance will be used to create arrays with compact designs without compromising performance. Arrays emulating biological systems will likely benefit greatly from the use of diffracting objects.

8.2 Helmet mounted array

The helmet mounted microphone array presented in Chapter 5 showed the same performance benefits experienced by mounting spherical arrays on diffracting spheres. While current applications have limited helmet arrays to counter-sniper systems, additional applications were presented in the tracking of land and air based vehicles. A localization algorithm based on the acoustic energy of environmental sound sources was presented and demonstrated using superdirective beamforming methods introduced in Chapter 4.

Results from tracking 3 AH-64 Apache helicopters demonstrated that using acoustic energy for source discrimination could be used to separate similar sources. Although the helicopters produced very similar acoustic signatures the localization method was able to track each helicopter as it powered up, taxied and took flight. When placed near a busy road the array reliably localized and tracked multiple cars approaching and departing from both directions. In a typical counter-sniper situation the array consistently determined the bearing of a shooter while establishing a range within 20% of the true value.

The results indicate that helmet arrays are not limited to localization applications involving impulsive acoustic events (e.g. gunshots). Sources with tonal characteristics resulting from turbines, rotors, engines, etc. were accurately localized and tracked in real-world environments. Although helmet arrays are viewed as suffering from diffractive

effects, the results show that accounting for these effects can actually produce reliable localization.

The military is constantly looking for technologies that will provide a benefit to our troops in combat. Applications that augment or enhance the performance of the human senses but can be used with little to no training are the most desireable, meaning that a consolidated effort between sensing and information presentation would be desirable. Pairing localization and classification algorithms will likely improve performance because spatial and acoustic characteristics will reinforce the confidence of detections.

8.3 Voice isolation using a headset mounted array

Chapter 6 presented a voice isolation algorithm implemented with a headset mounted array. A generalized sidelobe canceller (GSC) using both data independent beamforming and adaptive filtering methods was applied toward voice isolation. A headset mounted microphone array was placed on a manikin to simulate a real-world simulation on a human. A speaker mounted in place of the manikin's mouth was used to play voice signals while an external speaker played noise signals to act as an interferer. Two sets of experiments were performed using narrowband (tones spaced 200Hz apart between 200Hz and 4000Hz)and broadband noise (white noise between 20Hz and 4000Hz) signals to demonstrate the ability of the voice isolation algorithm to adapt. Within those two sets the external speaker was placed to the left of and directly in front of the manikin to examine the change in performance for different situations. Results from using a beamformer only and a beamformer coupled with an adaptive noise filtering algorithm were both presented.

For all experiment setups the "notch-out-noise" filtering outperformed the "beam-to-voice" filtering at suppressing noise. The performance of both filtering methods was worse for the cases when the interferer was located directly in front of the manikin. For narrowband noise signals the "beam-to-voice" filtering was only able to suppress a few tones while the "notch-out-noise" filtering nearly eliminated all tones (60dB in some

cases). Broadband results were similar with the "notch-out-noise" filtering suppressing noise by at least 15dB for most frequencies while the "beam-to-voice" filtering provided modest improvement.

The drastic improvement achieved in noise suppression when using the "notch-out-noise" filtering shows the disparity between simple beamforming and adaptive filtering methods. Specifically this indicates that using a data independent beamformer cannot match the performance of a statistically based adaptive algorithm when interfering noise sources are present.

Although not tested here, an additional benefit of the adaptive method is the ability to notch out multiple moving (with some caveats) noise sources whereas the data independent beamformer is fixed for all cases. This will lead to the development of a method that can identify multiple interferers, create multiple notches, and also adapt to situations where noise sources may not be constant (a few other voices, music, etc.). Also, the application of voice isolation coupled with voice recognition would doubtlessly prove a powerful tool in many acoustics applications.

8.4 Computational model of human head

A 3-D digital scan of a KEMAR manikin head wearing a set of headphones was presented in Chapter 2 as a tool for microphone array geometry optimization. The set of points was used in an equivalent source model to effectively compute the diffraction caused by the head and array for sounds coming from any direction. A sample geometry was chosen for a headset array prototype that was then used to experimentally verify model results, as discussed in Chapter 3. Several other prototypes were constructed and tested to further validate results, as seen in Chapter 7.

The results of the computational model were initially examined and validated up to 12kHz in section 3.4.4. Based on that validation the model was used to determine optimal array geometries for the application of transparent hearing. Subsequent prototype headset arrays were constructed, tested, and compared to model results. Section 7.4

shows again how the model is accurate by comparing model results to both the optimal array geometry and a raised microphone array geometry.

Results show that the computational model is an invaluable design tool for headset mounted arrays. However, there are some qualifications to these results. So long as arrays can be mounted without affecting the outside structure of the headset, the model should be accurate. If the array must modify the outside structure, small diffracting shapes with a maximum dimension on the order of 1 wavelength for the highest frequency of interest may not adversely affect the expected array response. Alternatively, so long as larger structures can be digitally modelled, the compuational model can be adapted accordingly. With all of these qualifications in mind, the computational model presented here produced phenomenal results and is currently a unique creation.

In addition to the headset array computational model presented in this thesis, several other digital scans were performed. The most intriguing of these scans was the KEMAR manikin head with pinnae inserts. Validation of this KEMAR model can be performed by computing the responses at the KEMAR ears and comparing those to the actual HRTF's. The intricate nature of the folds of the pinnae provides a challenge with respect to model complexity and subsequently, computation time. Additionally, while a general set of HRTF's can be measured from a number of sources, the true HRTF's for each person are unique. A couple of ideas come from this HRTF uniqueness: 1) a method for quickly determining a person's HRTF's and adjusting filtering methods appropriately or 2) a study of a wide range of HRTF's to determine subsets of head "shapes" that produce similar HRTF's. Both of these thoughts would expand the range of usefulnessof head mounted microphone arrays and their associated applications.

8.5 Microphone array geometry optimization method

With the specific application of transparent hearing as the target, the attributes of human hearing were used to create suitable optimization methods. Relying upon some aspects of duplex theory as the description of human hearing, a phase compen-

sation method was implemented for HRTF reconstruction. A cost function describing the match between the true human HRTF's and the recreated HRTF's was developed to assess the performance of different array geometries. Both an exhaustive and a shorter search routine revealed the optimized array geometry which was then constructed and verified experimentally. A second, modified geometry was constructed, verified experimentally, and then used in subjective testing with human subjects. The system, detailed in chapter 7, achieved its performance by acknowledging the limitations of human hearing to optimize microphone positions and tailor digital filters.

Applying the array optimization routine to the computational model results yielded several array designs. When the array geometries were sorted based on the cost function described in section 7.3.4, many geometries had similar characteristics. The first microphone (to which the HRTF was phase compensated) was the same for many of the best performing geometries while the remaining 2 microphones varied. Unfortunately the optimal array geometry did not take into account physical mounting restrictions based on internal hardware within the headset. Therefore a compromise was determined and implemented in the raised microphone headset array shown in section 7.4.2.

One finding from the results was that the position of the first microphone had the greatest impact on the recreation of HRTF's. Second and third microphones actually provide very small changes to more closely approximate HRTF's, but the gross match is left to the first microphone. The model provided an interesting insight that the optimal location for the first microphone is nearly in line with the ears, albeit on the outside of the headset. It is likely that this is affected by the set of source directions used in the optimization method, although in this case almost the entire upper hemisphere was used.

The implementation of a successful transparent hearing system with a single prototype provides the justification to pursue further improvements. As stated in section 7.4.2, the array prototype included microphone mounting structures not present in the computational model used for microphone position optimization. Future array prototypes may utilize MEMS microphones to more closely match the computational model and results.

From the standpoint of commercialization and marketability, MEMS microphones would also impact the overall appearance of headsets much less than raised microphone arrays.

8.6 Demonstration of transparent hearing system

A functional prototype using the modified optimal geometry was constructed and tested on human subjects undergoing psychoacoustic testing, as presented in Chapter 7. Transparent hearing strategies included an occluded control case, the Sennheiser TalkThrough system (2 microphones already equipped on the headset), a 3 microphone per headphone case using individualized filters, a 3 microphone per headphone case using generic filters, and a 1 microphone per headphone case using generic filters. The statistical significance of the improvement over other transparent hearing strategies was presented, proving that the optimization method afforded worthwhile gains.

Results of subjective testing were based on overall improvement and "cone of confusion" improvement as compared to an occluded condition. In the overall case only the 3 microphone individual and generic filter cases produced improvements that were statistically significant. The calculated p-values for these 2 methods were respectively 0.0029 and 0.0058, both less than the 0.05 level indicated by the 95% confidence interval. For the "cone of confusion" case the Sennheiser TalkThrough system also produced statistically significant improvements. The p-values for the 3 microphone individual, generic, and the TalkThrough system were 0.0023, 0.006, and 0.0035, respectively.

The results of the psychoacoustic tests help validate all of the previous design and optimization that went into creating the array and filters. The optimized array and its associated optimized filtering method did indeed improve localization over the occluded case. Statistical significance does indicate improvement over the baseline condition, in this case occluded hearing performance, but it does not definitively say which method was the best. The fact that both the 3 microphone individual and generic cases produced statistically significant improvement is a promising sign for the flexibility of filtering targets.

More extensive testing would expand this sample size to improve the weight of its results. Also, the testing of the most current competing transparent hearing systems would provide a basis for possible tradeoffs. Given the nature of human perception, subjective testing results with these other systems would also indicate directions for further improvements. Systems may claim to be the most accurate, but if the people wearing those systems cannot stand to wear them the accuracy may be moot. Deploying human subjects with wearable systems may be the most useful exercise as a laboratory environment is not typically able to recreate all of the dynamics of a real world environment. Lastly, while the 3 microphone individual filter case did produce the best p-values for both tests, the 3 microphone generic filter case also did well, warranting further examination of whether individual HRTF measurements are necessary.

APPENDIX A

Equivalent Source Method

As an initial approximation of the human head, the axes of a sphere can be adjusted to make it narrower in the x-dimension, a prolate spheroid. The general equation for a spheroid is

$$\frac{(x-x_0)^2}{a^2}+\frac{(y-y_0)^2}{b^2}+\frac{(z-z_0)^2}{c^2}=1 \tag{A.1}$$

where (x_0, y_0, z_0) is the center of the spheroid and (a, b, c) are the lengths of the major and minor axes. For a sphere all of the axes have the same length, i.e.

$$a=b=c=r \tag{A.2}$$

where r is the radius of the sphere. For a prolate spheroid 2 of the axes are equal in length and the third axis is shorter,

$$b=c \quad a<b \tag{A.3}$$

APPENDIX B

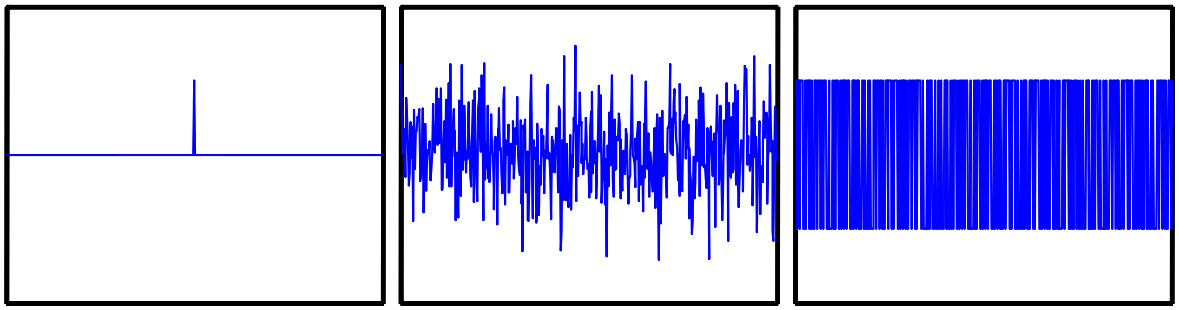
Characterization Stimuli

The simplest broadband signal is a perfect impulse or a series of perfect impulses. An impulse, by definition, has a flat frequency response, sufficiently exciting all frequencies equally. However, playing a pure impulse through a real speaker can introduce non-linearities when played at high amplitudes, and even at the maximum level, for nearly every speaker, the output of the speaker will not be enough to achieve a high SNR for a good measurement. A set of impulses can be used to average multiple measurements, but this method requires sufficient time between impulses to allow any reverberations present to decay. Another type of broadband signal used for characterization is Gaussian white noise. Unlike the case of impulses, the frequency response of a block of Gaussian white noise will not necessarily have a completely flat frequency response. That is to say that some frequencies will be represented with acceptable amplitudes, but some will have very small amplitudes, preventing accurate characterization. The solution to this frequency response issue is to perform characterization over a long enough time to insure sufficient average energy at every frequency of interest. One of the most intriguing broadband signals used for characterization is a set of Golay complementary codes. These codes, or pair of equal length signals, are completely comprised of elements equal to either +1 or -1 and have the property that the sum of their autocorrelations is equal to zeros for all delays except for zero [100]. A beneficial property of Golay codes is that all frequencies are guaranteed to have no more than 3dB more energy than the median energy of all frequencies. Also, the energy present in a set of Golay codes, in dB, is directly related to

the length of the codes, L, by

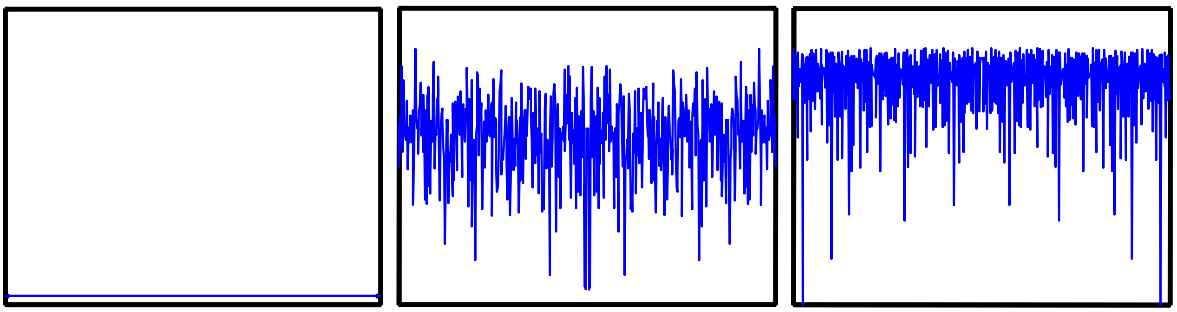
$$10 \log 2L \tag{B.1}$$

This is to say that if the SNR of a given set of Golay codes is not sufficient, all that is necessary is to increase the length of the Golay codes. Since the energy present is based on length, multiple stimuli and long averaging times are not necessary. A comparison of the time domain representations of all stimuli discussed can be seen in Figure B.1. The comparison is made for a block of data 512 samples in length for a single impulse, white Gaussian noise, and one Golay code from a pair of complementary codes. The frequency response of each code is shown in Figure B.2. The nature of Golay codes



(a) Single impulse (b) White noise (c) Golay code

Figure B.1 Sample reference stimuli for characterization.



(a) Single impulse (b) White noise (c) Golay code

Figure B.2 Frequency responses of reference stimuli for characterization.

makes them very attractive for tests where the measurement time and signal magnitude are concerns. Testing performed using human subjects to characterize the frequency and

spatial responses within the ear canal to sound sources in the far-field dictate that the sound level does not cause damage to human hearing. For measurements with arrays mounted on inanimate objects the sound pressure level may not be a concern during experiments, but with human subjects the goal is typically not to deafen the subjects. Small movements of human subjects may also disrupt the array manifold measurements if long sample times are used, so again Golay codes are an optimal solution [101]. Given a pair of complementary Golay codes, a and b, and the aforementioned autocorrelation property of said codes, only the original codes and 2 measurements are necessary to calculate the array manifold vector. In the frequency domain the codes are represented by A and B, and the responses of a microphone to those codes are AH and BH, where H is the linear transfer function between the sound source and the microphone. The original transfer function H is determined by

$$[AH]A^H + [BH]B^H = [AA^H + BB^H]H = 2LH \tag{B.2}$$

Although not shown here, so long as all physical equipment carrying signals to or from the experiment can be treated as having linear responses, those responses will also appear as a result of the previous calculation. Since those responses are treated as constant across all tests, their effect can be removed once the entire array manifold has been measured.

REFERENCES

[1] Brush Mountain Technologies, Inc. http://www.brushmttech.com.

[2] G. Duckworth, J. Barger, S. Carison, D. Gilbert, M. Knack, J. Korn, and R. Mullen, "Fixed and wearable acoustic counter-sniper systems for law enforcement," in *Proceedings of SPIE Conference on Sensors*, vol. 3577, pp. 210–230, 1998.

[3] G. Simon, M. Marti, . Ldeczi, G. Balogh, B. Kusy, A. Ndas, G. Pap, J. Sallai, and K. Frampton, "Sensor network-based countersniper system," in *Proceedings of the Second ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2004.

[4] P. Volgyesi, G. Balogh, A. Nadas, C. Nash, and A. Ldeczi, "Shooter localization and weapon classification with soldier-wearable networked sensors," in *Proceedings of the 5th International Conference on Mobile Systems, Applications, and Services*, pp. 113–126, 2007.

[5] M. Scanlon, "Helmet-mounted acoustic array for hostile fire detection and localization in an urban environment," in *Proceedings of SPIE Conference on Unattended Ground, Sea, and Air Sensor Technologies and Applications X*, vol. 6963, 2008.

[6] M. Takimoto, T. Nishino, K. Itou, and K. Takeda, "Evaluation of sound localization under conditions of covered ears," *Proceedings of Forum Acusticum 2005*, pp. L203–L206, 2005.

[7] T. Letowski and B. Amrein, "A08-T019 Phase I initial meeting," September 2008. ARL HRED Aberdeen Proving Grounds.

[8] A. Morton, D. Tack, D. Authority, and C. Boyne, "Literature review: Issues in restorative hearing," *Defence Research and Development Canada*, 2006.

[9] A. Bronkhorst, J. Verhave, and T. H. F. S. N. T. P. GROUP, "A microphone-array-based system for restoring sound localization with occluded ears," 2005.

[10] A. Goldstein, M. Johnson, W. Saunders, M. Vaudrey, and J. Carneal, "Natural hearing restoration part I: Theory," in *ACTIVE 2004*, INCE, 2004.

[11] M. Johnson, J. Carneal, and A. Goldstein, "Natural hearing restoration part II: Experimental results," in *ACTIVE 2004*, INCE, 2004.

[12] W. Chapin, P. Zurek, B. Shinn-Cunningham, N. Durlach, M. Bolas, and I. AUSIM, *Concept and technology exploration for transparent hearing*. Defense Technical Information Center, 2003.

[13] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[14] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[15] R. H. Inc., "Dictionary.com unabridged (v 1.1)," Mar 2009.

[16] D. Mennitt, *Multiarray passive acoustic localization and tracking*. PhD thesis, Virginia Polytechnic Institute and State University, November 2008.

[17] D. Mennitt and M. Johnson, "Multiarray passive acoustic source localization in urban environments," *Submitted to The Journal of the Acoustical Society of America*, 2008.

[18] R. Kennedy, D. Ward, and P. Abhayapala, "Nearfield beamforming using nearfield–farfield reciprocity," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997. ICASSP-97.*, vol. 5, pp. 3741–3744, 1997.

[19] A. Sleiman and A. Manikas, "Antenna array manifold: A simplified representation," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000. ICASSP-00*, vol. 5, pp. 3164–3167, 2000.

[20] A. Manikas, H. Karimi, and I. Dacos, "Study of the detection and resolution capabilities of a one-dimensional array of sensors by using differential geometry," *IEEE Proceedings-Radar, Sonar and Navigation*, vol. 141, no. 2, pp. 83–92, 1994.

[21] I. Dacos and A. Manikas, "Estimating the manifold parameters of one-dimensional arrays of sensors," *Journal of the Franklin Institute*, vol. 332, no. 3, pp. 307–332, 1995.

[22] A. Manikas, A. Sleiman, and I. Dacos, "Manifold studies of nonlinear antenna array geometries," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 497–506, 2001.

[23] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[24] P. Morse and K. Ingard, *Theoretical acoustics*. Princeton University Press, 1986.

[25] F. Wiener, "Sound diffraction by rigid spheres and circular cylinders," *The Journal of the Acoustical Society of America*, vol. 19, no. 3, pp. 444–451, 1947.

[26] F. Mechel, *Formulas of acoustics*. Springer New York, 2002.

[27] J. Bowman, T. Senior, P. Uslenghi, and J. Asvestas, "Electromagnetic and acoustic scattering by simple shapes," *Hemisphere Publishers Corp., New York*, 1987.

[28] P. Gillett, M. Johnson, and J. Carneal, "Performance benefits of spherical diffracting arrays versus free field arrays," in *Invited Paper: 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5264–5267, 2008.

[29] J. Carneal, M. Johnson, and P. Gillett, "Comparison of a diffracting and a non-diffracting circular acoustic array," in *2006 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. IV–IV, 2006.

[30] M. Johnson, J. Carneal, and P. Gillett, "Comparison of a diffracting and a non-diffracting cylindrical microphone array," in *Proceedings of the Thirteenth International Congress on Sound and Vibration*, 2006.

[31] E. Hobson, "The theory of spherical and ellipsoidal harmonics," *Cambridge, England*, 1931.

[32] M. Johnson, S. Elliott, K. Baek, and J. Garcia-Bonito, "An equivalent source technique for calculating the sound field inside an enclosure containing scattering objects," *The Journal of the Acoustical Society of America*, vol. 104, no. 3, pp. 1221–1231, 1998.

[33] M. Ochmann, "The source simulation technique for acoustic radiation problems," *Acustica*, vol. 81, no. 6, pp. 512–527, 1995.

[34] L. Rayleigh, "On the acoustic shadow of a sphere," *Philosophical Transactions of the Royal Society of London, Series A*, vol. 203, pp. 87–110, 1904.

[35] R. Duda, C. Avendano, and V. Algazi, "An adaptable ellipsoidal head model for the interaural timedifference," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. ICASSP-99*, vol. 2, pp. 965–968, 1999.

[36] T. Abhayapala and D. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002. ICASSP-02*, vol. 2, pp. 1949–1952, 2002.

[37] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.

[38] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 3094–3103, 2005.

[39] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, 2004.

[40] J. Meyer, "Beamforming for a circular microphone array mounted on spherically shaped objects," *The Journal of the Acoustical Society of America*, vol. 109, no. 1, pp. 185–193, 2001.

[41] R. Duda and W. Martens, "Range dependence of the response of a spherical head model," *The Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, 1998.

[42] G. Kuhn, "Model for the interaural time differences in the azimuthal plane," *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977.

[43] S. Laugesen, K. Rasmussen, and T. Christiansen, "Design of a microphone array for headsets," in *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pp. 37–40, 2003.

[44] J. Meyer, "Microphone array for hearing aids taking into account thescattering of the head," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on.*, pp. 27–30, 2001.

[45] V. Algazi, R. Duda, R. Duraiswami, N. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2053–2064, 2002.

[46] P. Gillett, M. Johnson, and J. Carneal, "A computational model for optimizing microphone placement on headset mounted arrays," in *Invited Paper: 123rd Convention of the Audio Engineering Society*, 2007.

[47] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[48] E. Brookner, "Phased-array radars," *Scientific American*, vol. 252, no. 2, pp. 94–102, 1985.

[49] E. Brookner, "Large phased-array radars," in *AIP Conference Proceedings*, vol. 178, pp. 165–205, 1988.

[50] W. Knight, R. Pridham, and S. Kay, "Digital signal processing for sonar," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1451–1506, 1981.

[51] R. Ward, D.B. Kennedy and R. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1023–1034, 1995.

[52] J. Harris, C. Pu, and J. Principe, "A monaural cue sound localizer," *Analog Integrated Circuits and Signal Processing*, vol. 23, no. 2, pp. 163–172, 2000.

[53] C. Pu, G. Harris, and J. Principe, "A neuromorphic microphone for sound localization," in *IEEE Conf. on Systems, Man, and Cybernetics*, vol. 2, pp. 1469–1474, 1997.

[54] S. Schelkunoff, "A mathematical theory of linear arrays," *Bell System Technical Journal*, vol. 22, no. 1, pp. 80–107, 1943.

[55] L. Kinsler, A. Frey, A. Coppens, and J. Sanders, "Fundamentals of acoustics," 1982.

[56] L. Rayleigh, "On our perception of the direction of a source of sound," *Journal of the Royal Musical Association*, vol. 2, no. 1, pp. 75–84, 1875.

[57] D. Ward, R. Kennedy, and R. Williamson, "Constant directivity beamforming," *Microphone arrays: signal processing techniques and applications*, pp. 2–17, 2001.

[58] A. Ishimaru, "Theory of unequally-spaced arrays," *Antennas and Propagation, IRE Transactions on*, vol. 10, no. 6, pp. 691–702, 1962.

[59] D. King, R. Packard, and R. Thomas, "Unequally-spaced, broad-band antenna arrays," *Antennas and Propagation, IRE Transactions on*, vol. 8, no. 4, pp. 380–385, 1960.

[60] J. Doles III and F. Benedict, "Broad-band array design using the asymptotic theory of unequallyspaced arrays," *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 1, pp. 27–33, 1988.

[61] J. Flanagan, J. Johnston, R. Zahn, and G. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *The Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1508–1518, 1985.

[62] T. Mueller, *Aeroacoustic measurements*. Springer, 2002.

[63] P. Ravetta, R. Burdisso, and W. Ng, "Noise source localization and optimization of phased array results (LORE)," *AIAA Paper*, vol. 2713, 2006.

[64] L. Brusniak, J. Underbrink, and R. Stoker, "Acoustic imaging of aircraft noise sources using large aperture phased arrays," *AIAA Paper*, vol. 2715, 2006.

[65] E. Gilbert and S. Morgan, "Optimum design of directive antenna arrays subject to random variations," *Bell Syst. Tech. J*, vol. 34, pp. 637–663, 1955.

[66] H. Cox, R. Zeskind, and T. Kooij, "Practical supergain," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 3, pp. 393–398, 1986.

[67] N. Owsley, "Enhanced minimum variance beamforming," 1988.

[68] O. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.

[69] J. Bitzer and K. Simmer, "Superdirective microphone arrays," *Microphone Arrays: Signal Processing Techniques and Applications*, pp. 19–38, 2001.

[70] H. Cox and R. Zeskind, "Reduced variance distortionless response (RVDR) performance withsignal mismatch," in *Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on*, vol. 2, pp. 825–829, 1991.

[71] M. Johnson, "Shielded microphone arrays for noise reduction," in *Proceedings of the 38th International Congress and Exhibition on Noise Control Engineering (Inter-Noise)*, 2009.

[72] B. Widrow and M. Hoff, "Adaptive switching circuits," *1960 IRE WESCON Convention*, vol. 4, pp. 96–104, 1960.

[73] T. Moon and W. Stirling, *Mathematical methods and algorithms for signal processing*. Prentice hall Upper Saddle River, NJ, 2000.

[74] D. Photiadis, "The relationship of singular value decomposition to wave-vector filtering in sound radiation problems," *The Journal of the Acoustical Society of America*, vol. 88, pp. 1152–1159, 1990.

[75] K. Cunefare, M. Currey, M. Johnson, and S. Elliott, "The radiation efficiency grouping of free-space acoustic radiation modes," *The Journal of the Acoustical Society of America*, vol. 109, no. 1, pp. 203–215, 2001.

[76] G. Daigle, M. Stinson, and J. Ryan, "Beamforming with air-coupled surface waves around a sphere and circular cylinder," *The Journal of the Acoustical Society of America*, vol. 117, no. 6, pp. 3373–3376, 2005.

[77] D. Mennitt, M. Johnson, and J. Carneal, "Bearing fusion methodologies for acoustic localization with networked microphone arrays," *Submitted to the Journal on Applied Signal Processing*, 2008.

[78] D. Mennitt, P. Gillett, C. J., and M. Johnson, "Noise source tracking using multiple diffracting arrays," in *ACTIVE*, 2006.

[79] D. Mennitt, P. Gillett, J. Carneal, and M. Johnson, "Tracking noise sources using multiple mobile microphone arrays," in *Proceedings of the Thirteenth International Congress on Sound and Vibration*, 2006.

[80] Z. Liu, M. Seltzer, A. Acero, I. Tashev, Z. Zhang, and M. Sinclair, "A compact multi-sensor headset for hands-free communication," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005*, pp. 138–141, 2005.

[81] Starkey Laboratories Inc. http://www.starkey.com/.

[82] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.

[83] F. Wiener, "The diffraction of a progressive sound wave by the human head," *The Journal of the Acoustical Society of America*, vol. 19, p. 290, 1947.

[84] T. Sandel, D. Teas, W. Feddersen, and L. Jeffress, "Localization of sound from single and paired sources," *The Journal of the Acoustical Society of America*, vol. 27, pp. 842–852, 1955.

[85] W. Feddersen, T. Sandel, D. Teas, and L. Jeffress, "Localization of high-frequency tones," *The Journal of the Acoustical Society of America*, vol. 29, pp. 988–991, 1957.

[86] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*. MIT press, 1997.

[87] J. Angell and W. Fite, "Monaural localization of sound," *Psychological Review*, vol. 8, no. 3, pp. 225–246, 1901.

[88] J. Angell and W. Fite, "Further observations on the monaural localization of sound," *Psychological Review*, vol. 8, no. 5, pp. 449–458, 1901.

[89] D. Batteau, "The role of the pinna in human localization," *Proceedings of the Royal Society of London. Series B, Biological Sciences*, pp. 158–180, 1967.

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

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

[92] W. Hartmann, "How we localize sound," *Physics today*, vol. 52, no. 11, pp. 24–29, 1999.

[93] H. Wallach, "On sound localization," *The Journal of the Acoustical Society of America*, vol. 10, pp. 270–274, 1939.

[94] F. Wightman and D. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, vol. 105, pp. 2841–2853, 1999.

[95] H. Wallach, E. Newman, and M. Rosenzweig, "The precedence effect in sound localization," *American Journal of Psychology*, vol. 62, no. 3, pp. 315–336, 1949.

[96] S. Smith, *The scientists and engineers guide to digital signal processing*. California Technical Publishing, 1999.

[97] R. Litovsky, H. Colburn, W. Yost, and S. Guzman, "The precedence effect," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, 1999.

[98] M. Burkhard and R. Sachs, "Anthropometric manikin for acoustic research," *The Journal of the Acoustical Society of America*, vol. 58, no. 1, pp. 214–222, 1975.

[99] M. Johnson, P. Gillett, E. Perini, A. Toso, and D. Harris, "Microphone array optimization for a hearing restoration headset," in *127th Convention of the Audio Engineering Society*, 2009.

[100] M. Golay, "Complementary series," *Information Theory, IRE Transactions on*, vol. 7, no. 2, pp. 82–87, 1961.

[101] B. Zhou, D. Green, and J. Middlebrooks, "Characterization of external ear impulse responses using Golay codes," *The Journal of the Acoustical Society of America*, vol. 92, no. 2, pp. 1169–1171, 1992.