# Design of e-textiles for acoustic applications

**Ravi R. Shenoy**

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

Master of Science
in
Electrical Engineering

Dr. Mark T. Jones, Chair
Dr. Thomas L. Martin
Dr. William H. Tranter

Bradley Department of Electrical and Computer Engineering
Blacksburg, Virginia

Keywords: e-textiles, Microphone array, Beamforming, DOA, Source Localization

# Design of e-textiles for acoustic applications

**Ravi R. Shenoy**

## Abstract

*The concept of replacing threads with flexible wires and sensors in a fabric to provide an underlying platform for integrating electronic components is known as e-textiles. This concept can be used to design applications involving different types of electronic components including sensors, digital signal processors, microcontrollers, color-changing fibers, and power sources. The adaptability of the textiles to the needs of the individual and the functionality of electronics can be integrated to provide unobtrusive, robust, and inexpensive clothing with novel features. This thesis focuses on the design of e-textiles for acoustic signal processing applications. This research examines challenges encountered when developing e-textile applications involving distributed arrays of microphones. A framework for designing such applications is presented. The design process and the performance analysis of two e-textiles, a large-scale beamforming fabric and a speech-processing vest, are presented.*

*I would like to dedicate this thesis to my grandfather*
*Late Kadri. Srinivas. Shenoy. I would not be*
*who I am and where I am without his blessings.*

# Acknowledgements

I wish to thank Dr. Mark T. Jones for accepting me as a research assistant and for serving as my advisor and guiding this research. I am highly indebted for his insightful advice and encouraging words which helped me meander through my graduate research.

I am extremely thankful to Dr. Thomas L. Martin for his help and guidance. My gratitude goes to him for his technical support and suggestions that made my work a success.

I would like to thank Dr. William H. Tranter for his support and motivation. I sincerely appreciate his willingness to serve on my committee.

Special thanks goes to Zahi S. Nakad, David L. Lehn and Dr Jae H. Park for their help during the research.

I would like to thank my friends who have made my stay at graduate school enjoyable and the people in my research group who have made every aspect of research exciting and interesting for me.

Finally, I would like to thank my family for their unconditional love and faith on me. Their persistent support and high expectations from me has driven me throughout my life. With their continued wishes and backing, I hope to live up to their expectations.

Ravi R. Shenoy

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

"Most people examine fabric swaths for texture and color; Maggie Orth checks them for voltage readings. Orth is the CEO of Cambridge, MA-based International Fashion Machines, a developer of electronic textiles in which fabrics act as electrical conduits, enabling data transfers within clothing" [1]. The mere fact that the fabric swath is being used as electrical conduits opens up a new dimension in which traditional fabric is being viewed. Textiles themselves provide the mechanical, aesthetic, and flexible qualities required for clothing. Clothing is the element that is almost always present and is customizable to each and every human being. Thus, if clothing can provide some intelligent features in a natural, robust, inexpensive, and unobtrusive way, then it can enhance the quality of already existing applications. New applications can also be developed to exploit such textiles. These intelligent features can then involve computational entities such as Digital Signal Processors (DSP) and microcontrollers, as well as distributed sensors, color-changing fibers, ultrasonic transducers, and optical fibers. In addition, the data and/or power-distributing conductors can be incorporated in a normal textile during the weaving process by replacing some of the threads with conducting wires. This generic concept of including intelligent features in fabrics has

been under research for almost over a decade now, under different names such as electronic textiles, e-textiles, smart textiles, wearable computers, and ubiquitous computers.

This thesis addresses the design issues for integrating acoustic applications on the e-textiles. A framework for designing two applications, a large-scale beamforming fabric (Chapter 4) and a speech-processing vest (Chapter 5), is described. The importance of simulation to explore the complete design space is emphasized and the performance of the prototypes developed for the two applications is reported.

## 1.1 Motivation

One of the initial successful prototypes developed by the VT e-textiles group was a glove for user input that employed piezoelectric transducers to sense the movement of fingers [2]. The glove demonstrated the seamless possibilities that e-textiles offered. This was the primary motivation for exploration into the field of e-textiles. Integrating wires into the textiles is the fundamental idea upon which e-textiles are based. Weaving a large fabric with some of the threads being replaced by flexible wires was the necessary condition that had to be applied and evaluated before building future more complex e-textile applications. It was necessary to evaluate e-textiles in the presence of long data paths, distributed sensors, energy sources, and signal-processing entities. These were the reasons for designing the large-scale beamforming fabric (Chapter 4). The goal of the application was to estimate the direction of arrival of a distant moving vehicle using a distributed array of microphones. In addition, location estimates for the moving vehicle were to be estimated using the direction of arrival estimates from more than one cluster.

Active research has been going on in the area of microphone arrays for speech enhancement and hands-free telephony [3]. The inherent spatial diversity existing in an array of micro-

phones can be used to retrieve high quality acoustic signals by overcoming the effects of reverberations and suppressing unwanted noise sources. For such applications, an accurate fix on the primary speaker, as well as the knowledge of other speakers and interfering sources, is necessary. Although microphone arrays look very promising for developing hearing aids and speech enhancement applications for conference rooms, an efficient way to incorporate microphone arrays on the human body is essential. Textiles can be used to provide an underlying platform for placing distributed sensors and processors. As data and power paths are existent in an e-textile, it can be used to host an array of distributed microphones. Thus, the usefulness of the application and the presence of an extremely suitable underlying platform for realizing it in a suitable form factor were the principle motivations for developing the speech-processing vest (Chapter 5). Figure 1.1 shows a possible scenario where microphone arrays can be used.

Both of these applications involved signal processing aspects for acoustic frequencies. As an e-textile was the underlying platform on which these applications were to be implemented, this thesis addresses the challenges encountered for such applications.

## 1.2 Contributions

This thesis designs and evaluates the performance of acoustic signal processing applications when they are integrated into an e-textile. The fundamental issues for the integration of electronics into textiles are described and addressed. Because the range of available options for variables such as the placement of sensors, the number of sensors, and hardware components are huge, simulations are required. The use of simulations to check the quality of the application designed under various physical environments is also discussed. The values for the design variables that give satisfactory results are used for building prototypes. The

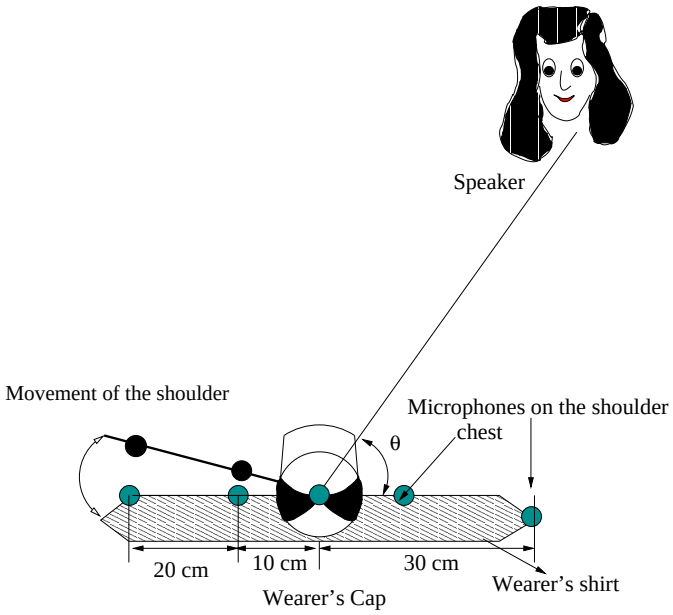


Figure 1.1: Speech-processing vest

results of the simulations and the tests conducted on the prototypes demonstrate that both simulation and prototyping are necessary for building e-textile applications. Thus, the thesis presents a complete framework for designing and building e-textiles for acoustic signal processing applications. An analysis of the large-scale beamformer as a function of number of microphones, placement of microphones, sampling rates, and signal-to-noise ratio is presented.

## 1.3 Thesis Organization

The thesis is organized in the following manner. Chapter 2 gives a background to the work undertaken in the thesis. Chapter 3 explains the fundamental issues involved in the integration of the electronics and textiles. The need for exploring the complete design space through the simulations and developing a limited number of prototypes is explained. It also

gives a description of the simulation environment and a framework for the design of e-textiles for acoustic applications. Chapter 4 describes the complete design process of a large-scale beamforming fabric that estimates the direction of arrival of moving vehicle in an outdoor environment. A description of the actual prototype and results from a field test are included in the chapter. Chapter 5 discusses the issues involved in the using microphone arrays on the human body. Two methods for direction of arrival using microphone arrays are considered. Results of the simulation for the performance of the two methods in the presence of shoulder movement are presented. The last two sections give a description of the prototype followed by the results of tests conducted in an auditorium and a teleconferencing room.

# Chapter 2

# Background

The first section of the chapter gives a general overview of the nature of wearable computer applications. The second half of the chapter gives a brief overview of the fundamentals of microphone arrays.

## 2.1 Applications

Over a decade of research has led to the development of a broad range of applications for wearable computers. These applications are in diverse areas including medicine, games, user interfaces (UI), context awareness, clothing for special environments, and devices to help people with disabilities.

One of the early prototypes in the field of medicine was a wearable device for physicians [4]. This device was designed as a means for reviewing medical literature during dermatological examinations. The device could receive and transmit digital information in the form of text and images. This information was then used to create patient records and provide remote

consultations. This prototype demonstrated the usefulness of wearable computers in the field of medicine. The sensate liner project designed a garment that contained a mesh of optical and electrical fibers woven into a normal textile [5]. The output signals from these sensing elements were connected to a block consisting of a processor and a transmitter. This integrated garment was used to monitor the medical condition of a patient.

The wearable motherboard from Georgia Tech provides an extremely versatile framework for incorporating sensors, monitoring devices, and processing devices in an unobtrusive manner [6]. It has embedded sensors and a network bus to carry information to and from the sensors. The application was designed as a combat casualty care fabric that was able to detect bullet wounds and vital body signs during combat operations.

Wearable computers can be of great help in assisting the visually challenged to sense an approaching object. A device called "VibroTach" was built to convey the velocity of the moving object to the wearer [7]. A doppler motor was used to convert the velocity of the approaching object into the proportional spin of the motor. The vibrations arising due to the spin were felt by the wearer and used to feel the velocity of the object. A similar application was developed using pyroelectric and ultrasound transducers under the "People Sensor" project [8]. This application could, in the near field, differentiate between an animate (human) and an inanimate approaching object. Sensors for this device were clasped onto the wearer's shirt or jacket. The output was transmitted to the user through a device on his/her waist.

In the field of entertainment, wearable computers are finding a place in virtual reality and gaming applications. *Game-City* is one such gaming application that has developed a multi-interface mixed reality game [9]. It uses wearable computer interaction and a mixed reality computer environment to incorporate fantasy features into already existing games. Sensors such as microphones and accelerometers are used in other games [10].

Garments with sensors embedded in them can be of great help in adapting to harsh physical environments. One such garment is a smart fabric for arctic environments [11]. This garment has attached GSM and GPS devices, heart rate monitoring system, small display unit, sensors to detect the impact of falling, and underwear that is able to regulate body temperature. All of the devices, except the body temperature regulation device, were fitted onto an outer jacket. Another project focused on a smart heating fabric [12]. This fabric had embedded temperature sensors, humidity sensors, and a palm top for display. The heating element consisted of twelve resistive carbon-heating panels. A microprocessor was used to gather all of the signals and regulate the heating unit. The power supply was carried on a separate belt instead of being mounted on the undershirt.

A speech enhancement vest was developed at MIT Media Laboratories [13]. This fabric consisted of three microphones. The signals from these microphones were processed by a personal computer. The goal of the design was to localize an active speaker and enhance the quality of the speech by processing the signals obtained from multiple microphones. The application was able to distinguish between the commands given by a user and nearby speech from a speaker.

## 2.2 Issues Addressed

The challenges encountered while designing the applications mentioned in Section 2.1 have been the principle motivation for focused research in some of the areas. Research has been done to look at the placement of different types of sensors including microphones, temperature sensors, humidity sensors, sensors to determine rate of heartbeat, and impact sensors [5] [6] [12]. Focus on the fusion of sensor signals instead of algorithms was also done in [14]. A project on electric suspenders demonstrated the utility of designing a single

system-level bus within the wearable fabric for information exchange and power distribution [15]. It focused on using conductive webbings (durable conductors directly in the matrix of the weave) for integrating the electronics on conventional clothing. A single battery source for all of the devices and sharing of information between different types of devices was shown to be possible because of existence of such a bus. Investigation was done to look for possible locations on the body for placing text input and pointing devices [16]. The efficiency and the usability of three different types of keyboard namely, forearm keyboard, virtual keyboard, and kordic keypad, were investigated in [17].

System level analysis of power consumption was the central focus for the real-time speech translator application [18]. It analyzed the performance and the power consumption of the wearable application as a function of processor speed, memory size, and the type of secondary storage.

A small badge was designed to capture the interaction of children with their surroundings [19]. This badge could fit inside a shirt pocket. The hardware consisted of sensors, a DSP, and a radio. The data was communicated through a middleware infrastructure using a wireless link for user-friendly data access. This project presented the challenges involved in the design of hardware and its integration with a wireless link for data transfer.

Lastly, comfort level assessment for wearable computers was done in [20]. A tool was presented to assign comfort levels for applications. The ratings were based on six dimensions, namely emotions, attachment, harm, perceived change, movement, and anxiety.

Thus, efforts have been made to analyze issues such as power consumption, placement of sensors, placement of devices, comfort assessment for wearers, batteries, system level architecture, software, and networking.

## 2.3 View of e-textiles

Wearable computer is a term used to describe any integrated device or a network of distributed electronic components placed on the human body. One of the sub-classes of this broad field is e-textiles. The design of an application involves issues relating to both the textile as well as those of the electronics industry.

The basic difference between an e-textile and other wearable computer applications is the presence of data and/or power distributing conductor's woven into an e-textile. Most of the wearable computer applications are stand-alone devices attached at specific locations on the body. Thus, the focus of wearable computers is on the development of applications or accessories. On the other hand, e-textile research focuses on providing an underlying platform, which can then be used to mount various devices or sensors. Hence the issues faced in the development of an e-textile are very different from the issues seen in general wearable computing application design.

## 2.4 Microphone Arrays

Microphone arrays are being used extensively in speech enhancement, source localization, source separation, speech-to-text translation, and speech recognition applications. In a single source case, phase shifting and adding operation of signals from all of the microphones results in signal-to-noise ratio improvement [21]. The improvement achieved is a factor of the number of microphones present in the array. The same idea can be used in source localization applications, where the time difference between signals received at different microphones can be used to estimate the location of the source. The output of a microphone array can also be steered to get a better signal quality in a desired direction while at the same time

suppressing signals from unwanted directions. The process of steering the response of the array in a particular direction is known as beamforming [21]. Source separation applications depend upon the beamforming principle to separate out more than one active source at any given time. Most of the speech-to-text and speech recognition applications require a very high signal-to-noise ratio. Microphone arrays can thus directly help in improving the quality of signal when the source is at a far away distance and/or the quality of the speech is degraded by room acoustics. The basic principle of estimating the location of single source is explained in the following section.

The reference or the origin of the coordinate system is fixed at $M_0$. As shown in Figure 2.1, an acoustic source located at $\{\rho_0, \theta_0\}$ (polar coordinates) is assumed to be emitting a perfect sinusoid $s(t)$ of frequency $f$ Hz. $c$ is the velocity of sound in the atmosphere and its value depends on the atmospheric conditions such as humidity and temperature. For our purposes, its value is assumed to be a constant at 342 m/sec. Microphone $M_1$ is separated from $M_0$ by a distance $d$. The signals received at microphones $M_0$ and $M_1$ are attenuated and phase delayed versions of the radiated signal. The distance between the microphone $M_1$ and the speaker can be calculated as

$$\rho_1 = \sqrt{(\rho_0 * sin(\theta_0))^2 + (d + \rho_0 * cos(\theta_0))^2} \tag{2.1}$$

In a noise-free environment without any multipaths, both attenuation and phase delay are functions of the distance traveled by the wave. Attenuation is typically assumed proportional to -2 power of the distance traveled by the wave [22]. Because the distance between the microphones is assumed to be much smaller than the distance between the source and the microphone array, the signal strengths are assumed to be equal at the microphones. The resultant phase $\phi_i (i = 0, 1)$ seen at any one of the microphones with respect to the source

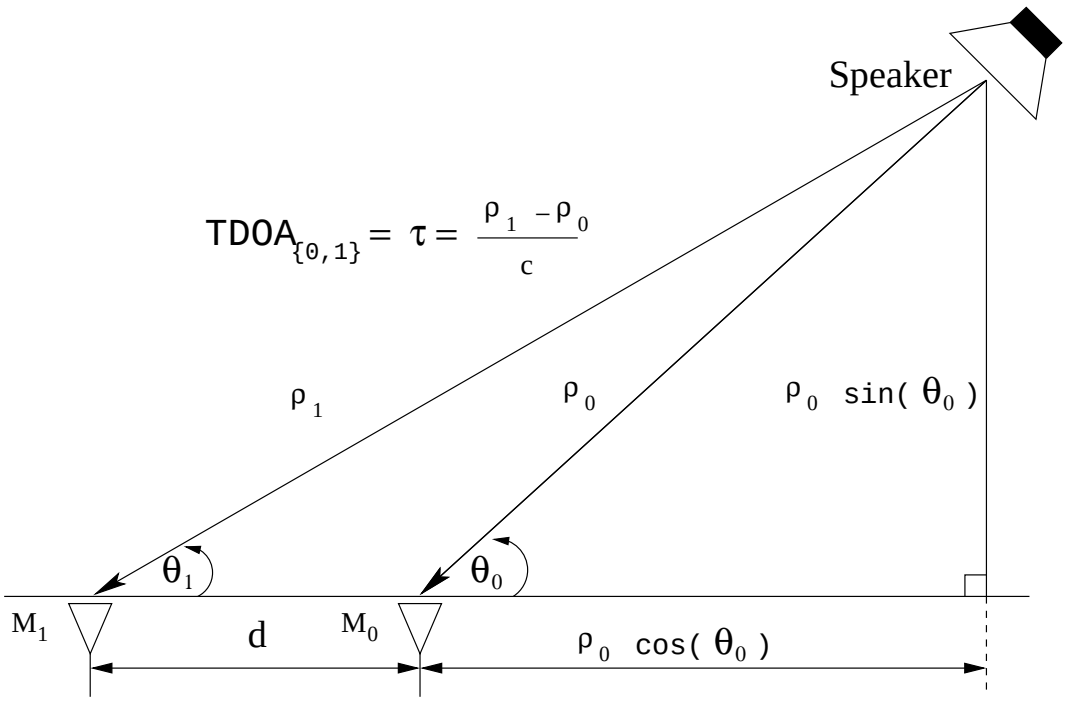


Figure 2.1: Time difference of arrival (TDOA) between a pair of microphones

signal is given by

$$\phi_i = \frac{2\pi\rho_i}{\lambda} \tag{2.2}$$

where $\lambda = c/f$ wavelength of the sinusoidal wave in the atmosphere. Because the wavefront has to travel some extra distance $(\rho_1 - \rho_0)$, the signal seen at $M_1$ will be lagging in phase with respect to the signal at $M_0$. The time difference of arrival between the signals arriving at $M_0$ and $M_1$ is given by

$$\tau_{0,1} = \frac{(\rho_1 - \rho_0)}{c}. \tag{2.3}$$

An acoustic source is said to be in the far field when the distance between the source and the microphone array is $\rho_0 > 2d^2/\lambda$ [23]. Under far field conditions, the wave incident on the microphone array is a plane wave. Figure 2.2 shows such a plane wave incident on a uniform linear array consisting of 4 microphones $\{M_0, M_1, M_2, M_3\}$, where the distance between the adjacent microphones is $d$. Because of the far field assumption, the path difference and the

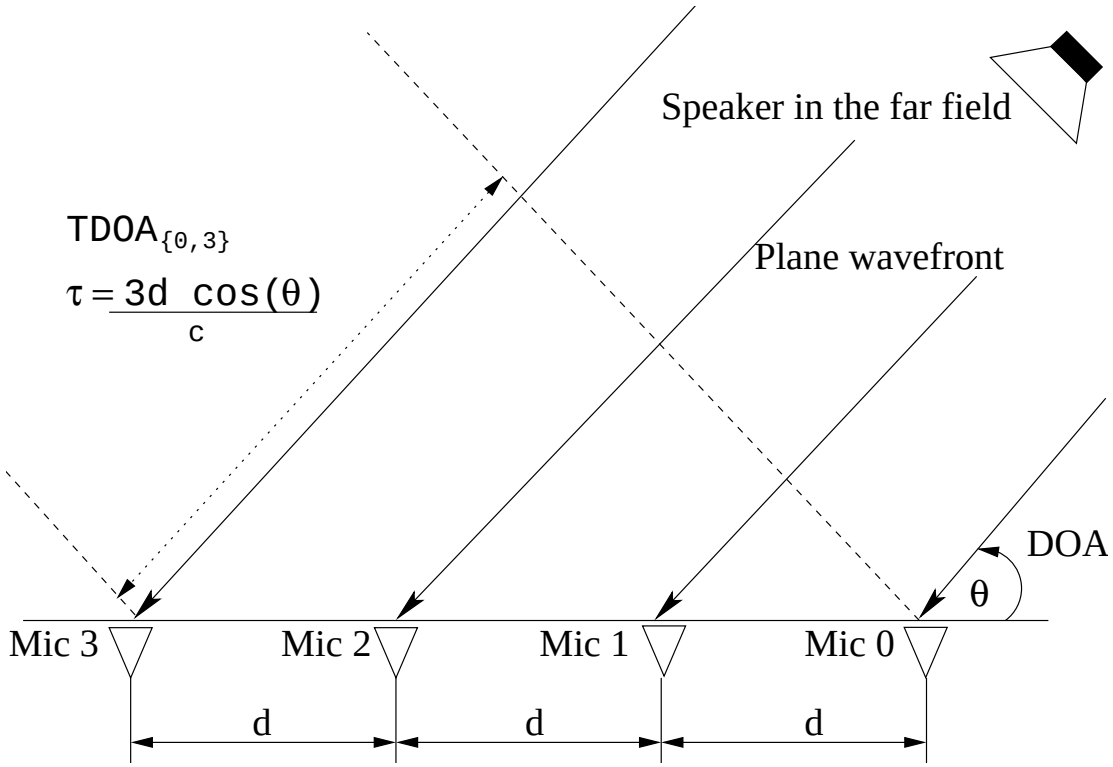


Figure 2.2: Uniform Linear Array

time difference of arrival between a pair of microphones $i$ and $j$ can be approximated as

$$\begin{aligned} \Delta\rho_{ij} &= \rho_j - \rho_i \\ &\approx d_{ij}\ cos(\theta) \end{aligned} \tag{2.4}$$

$$\begin{aligned} \tau_{ij} &= \frac{\Delta\rho_{ij}}{c} \\ &\approx \frac{d_{ij}\ cos(\theta)}{c} \end{aligned} \tag{2.5}$$

and the corresponding phase difference is given by

$$\begin{aligned} \Delta\phi_{ij} &= 2\pi f \tau_{ij} = \omega\ \tau_{ij} \\ &= \frac{2\pi d_{ij}\ cos(\theta)}{\lambda}. \end{aligned} \tag{2.6}$$

As shown in the following equations, the signals seen at the microphones $\{M_1, M_2, M_3\}$ are the delayed versions of the signal received at $M_0$. The delays at microphones $M_1$, $M_2$, and

$M_3$ with respect to $M_0$ are in ascending order as

$$\begin{aligned} x_0(t) &= as(t - \rho_0/c), \\ x_1(t) &= x_0(t - \tau_{01}) = x_0(t - d\,cos(\theta)/c), \\ x_2(t) &= x_0(t - \tau_{02}) = x_0(t - 2d\,cos(\theta)/c), \qquad \text{and} \\ x_3(t) &= x_0(t - \tau_{03}) = x_0(t - 3d\,cos(\theta)/c), \end{aligned} \tag{2.7}$$

where $a$ is the attenuation factor. If the angle of incidence $\theta$ is greater than $\pi/2$, then the signals received at microphones $M_3$, $M_2$, and $M_1$ are leading in phase as compared to signal at $M_0$.

## 2.5 Spatial Aliasing

There is a restriction on the maximum distance by which the two adjacent microphones can be separated. This restriction arises because phases greater than $2\pi$ are wrapped around such that phase $\phi$ is $-\pi \leq \phi \leq \pi$. The wavelength of an acoustic wave is $\lambda = c/f$. For a given angle of incidence $\theta$ and separation $d$, the phase difference $\phi$ between two microphones is given by $2\pi dcos(\theta)/\lambda$ (2.6). Thus, given a phase difference, the corresponding angle of incidence can be estimated as

$$\theta = \arccos(\phi\ \frac{\lambda}{2\pi\ d}). \tag{2.8}$$

As shown in Figure 2.3, the time difference of arrival for two signals with angles of incidence $\theta$ and $-\theta$, i.e. one coming from the front and the other coming from the back, will be same. In order to maintain one-to-one correspondence between $\theta$ and time difference of arrival, only one side of the array is considered such that $0 \leq \theta \leq \pi$. A unique relation between $\theta$

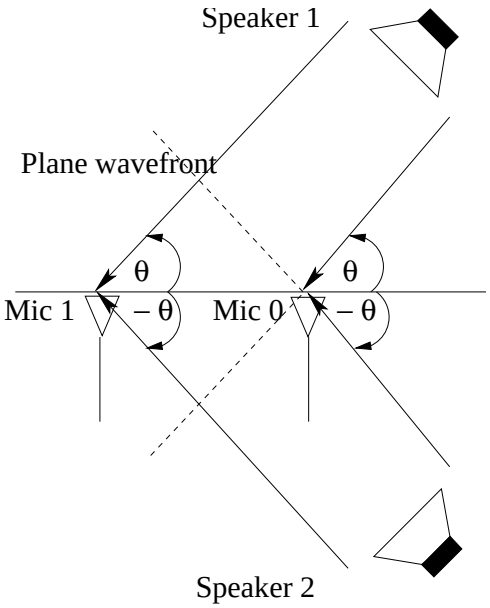


Figure 2.3: Front-back ambiguity

and $\phi$ can only be maintained by satisfying both the conditions

$$0 \leq \theta \leq \pi, \qquad \text{and} \tag{2.9}$$

$$-\pi \leq \phi \leq \pi. \tag{2.10}$$

Both of these conditions can only be satisfied by having

$$d_{max} \leq \lambda/2. \tag{2.11}$$

To illustrate this point further, Figure 2.4 shows the condition where the phase difference between two microphones for different angles of incidence are identical when $\lambda/2 < d$. Hence these distinct incident angles are indistinguishable.

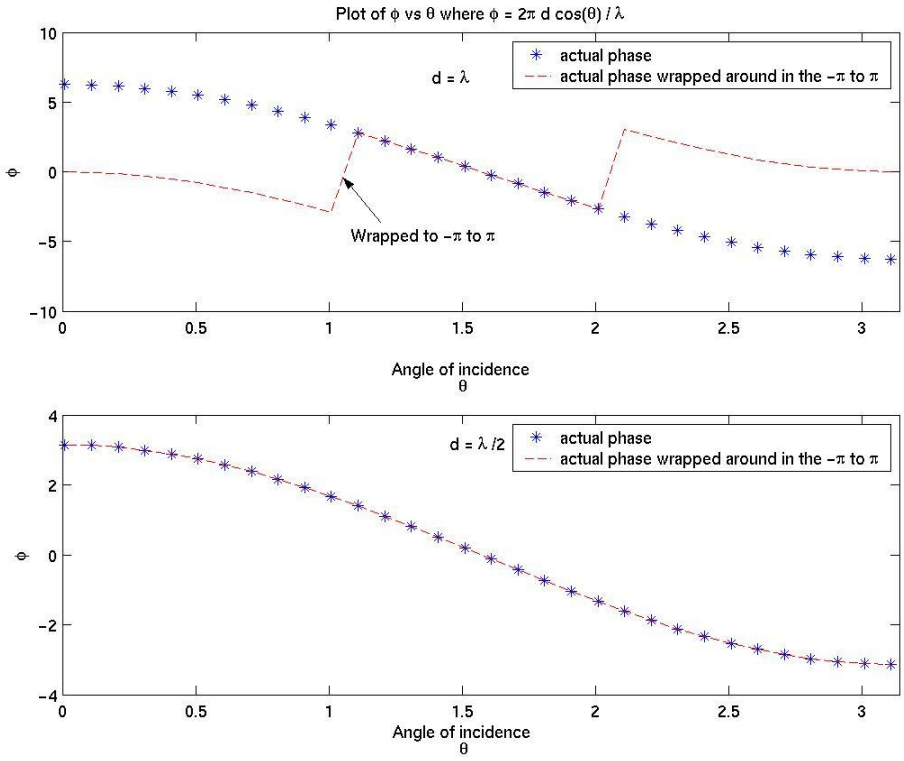


Figure 2.4: The plot on top shows ambiguity due to $d > \lambda/2$. The bottom plot does not have this phase ambiguity

## 2.6 Direction of Arrival Estimation

A uniform linear array with a source in the far field is used to explain the basic principles involved in direction of arrival estimation. As can be seen from the Equation (2.7), the signals received at the individual microphones are delayed versions of the same signal. This delay is a function of angle of incidence, $\theta$, and the separation between a pair of microphones, $d_{ij}$. A simple beamformer would shift and add the signals from all of the microphones to realign their phases and get a resultant in-phase stronger signal. Because the angle of incidence is unknown, a set of values for the delays for each possible angle of incidence is pre-calculated. Assuming a particular angle of incidence, the signals from all the microphones are advanced

by the corresponding delays for that angle of incidence. After advancing them in time, the signals from all of the microphones are then summed. The peak power for the summed signal indicates that the signals from all of the microphones are in-phase. The angle at which peak power is obtained is the estimated angle of incidence. This method of estimating the direction of arrival is known as the delay-and-sum beamformer [24],

$$\theta = \underset{\theta}{argmax} \sum_{j=0}^{M-1} s(t - \tau_{0j} + \Delta_{j,\theta}), \tag{2.12}$$

where

$$\Delta_{j,\theta} = \frac{d_{0j} cos(\theta)}{c}.$$

The same beamforming principle can be applied in the frequency domain. The signals received at the microphones can be written in vector notation as

$$\begin{aligned} x(t) &= \begin{bmatrix} x_0(t) & x_1(t) & \dots & x_{M-1}(t) \end{bmatrix}^T, \quad \text{and} \\ x(t) &= \begin{bmatrix} a_0 x_0(t) & a_1 x_0(t - \tau_{01}) & \dots & a_{M-1} x_0(t - \tau_{0M-1}) \end{bmatrix}^T, \end{aligned} \tag{2.13}$$

where $a_i$ and the $\tau_{0i}$ are the attenuation and phase delays of the $i^{th}$ microphone with respect to the $0^{th}$ microphone in the array. $\tau_{0i}$ is given by the equation (2.5) and attenuation is assumed equal for all of the microphones because of the far field assumption. $x(t)$ can be represented in the frequency domain as

$$\begin{aligned} X(\omega) &= \begin{bmatrix} X_0(\omega) & X_1(\omega) & \dots & X_{M-1}(\omega) \end{bmatrix}^T, \quad \text{and} \\ X(\omega) &= \begin{bmatrix} a_0 X_0(\omega) & a_1 X_0(\omega) e^{-j\omega\tau_{01}} & \dots & a_{M-1} X_0(\omega) e^{-j\omega\tau_{0M-1}} \end{bmatrix}^T. \end{aligned} \tag{2.14}$$

This can be broken down into

$$\begin{aligned} X(\omega) &= X_0(\omega) \begin{bmatrix} 1 & a_1 e^{-j\omega\tau_{01}} & \dots & a_{M-1} e^{-j\omega\tau_{0M-1}} \end{bmatrix}^T, \quad \text{and} \\ X(\omega) &= X_0(\omega) A(\omega). \end{aligned} \tag{2.15}$$

The signals from all of the microphones are weighted by complex coefficients and the angle that gives the peak power for the resultant signal is the estimated direction of arrival [24].

$$y(t) = \sum_{j=0}^{M-1} w^*(\omega, \theta) x(t), \quad \text{and} \tag{2.16}$$

$$w(\omega, \theta) = \begin{bmatrix} w_0(\omega, \theta) & w_1(\omega, \theta) & \dots & w_{M-1}(\omega, \theta) \end{bmatrix}^T. \tag{2.17}$$

The Equation (2.16) can be written in frequency domain as

$$Y(\omega) = W^H(\omega, \theta) X(\omega). \tag{2.18}$$

The power of the resultant signal is given by

$$\begin{aligned} P(\omega, \theta) &= \Big| y(t) \Big|^2 \\ &= Y(\omega) Y^*(\omega) \\ &= \Big( W^H(\omega, \theta) X(\omega) \Big) \Big( W^H(\omega, \theta) X(\omega) \Big)^H \\ &= \Big( W^H(\omega, \theta) X(\omega) \Big) \Big( X^H(\omega) W(\omega, \theta) \Big) \\ &= W^H(\omega, \theta) \Big( X(\omega) X^H(\omega) \Big) W(\omega, \theta) \\ &= W^H(\omega, \theta) R W(\omega, \theta), \end{aligned} \tag{2.19}$$

where R is the covariance matrix of the input signal $R = E\{x(t)x^T(t)\}$ [25]. It can be shown that, when a simple phase shifting and adding operation in time domain is to be carried out, the weight $W(\omega, \theta)$ is simply $A(\omega, \theta)$ [21]. Thus, the direction of arrival can be estimated as

$$\theta = \underset{\theta}{argmax}\ A(\omega, \theta)^H R A(\omega, \theta). \tag{2.20}$$

## 2.7 Speech Signals

Typical speech signals span a range of frequencies from 100 Hz to about 5000 Hz. This makes the signal very broadband [26]. Thus, the narrowband beamforming principles cannot be applied directly to find the direction of arrival. Speech can be broadly classified as voiced and unvoiced signals. In voiced speech, depending on its size, certain frequencies resonate within the vocal tract. The energy in these frequencies is increased when the reflected waves coincide and reinforce each other. These formant frequencies appear as humps in the spectrum and are known as formants. Voiced speech consists of many such formants, but only the first few are important in the analysis of the speech. The fundamental formants typically lie in the range of about 220 Hz to 1000 Hz and vary from person to person. In voiced speech, the energy in the frequency spectrum decays at the rate of 6 dB/octave [26]. In the case of unvoiced speech, there is no energy concentration and the spectrum is more or less flat until approximately 5 KHz.

## 2.8 Source Localization Strategies

The localization strategies for speech can be broadly classified into three categories, namely steered response power, high-resolution spectral beamformers, and those based on time delay estimates [3].

Steered response power (SRP) localization strategies weight the signals from different microphones and sum the signals to find the maximum power. The optimum maximum likelihood estimator uses this strategy to steer the output of the beamformer in various directions and find the maximum power. Although these types of beamformers can give optimal results, the real disadvantage of these beamformers is the computational complexity [27]. As in

the maximum likelihood estimator, the direction estimation involves solutions to nonlinear equations and this is the primary reason why these beamformers are not implemented in real time. In the presence of some *a priori* knowledge of the source signal and spectral content of the background noise, the complexity of the solution to the nonlinear problem can be reduced [28]. However, in most practical real-time speech applications, this information is not available.

The second category of beamformers is based on high-resolution spectral concepts such as autoregressive modeling, minimum variance, and other eigenvalue based techniques [24] [29] [30]. They use the signal covariance matrix for the direction of arrival estimation. In most practical situations, the signal covariance has to be estimated from observed data. These beamformers are designed for narrowband signals. These can be extended to wideband signals by processing each of subspace of the signal spectrum as a narrow band signal. These can also be used in multi-source scenarios. The disadvantage of these beamformers is that room reverberations tend to have severe performance degradations.

The third type of beamformer estimates the time difference of arrival to compute the direction of arrival. Techniques for estimating the time difference of arrival include generalized cross correlation (GCC) [31], GCC-PHAT, GCC-ML [32], and SRP-PHAT [3]. Because speech signals are interspersed by pauses, the time delay estimates have to be obtained from the short analysis interval of speech signals (typically 20 ms - 30 ms). The advantage of these beamformers is that they require little computation when compared to other methods. They are very well suited for real time implementation. This category of beamformers cannot be used for multi-source scenarios. They can be used, however, to localize one of the many possible active sources at a given instant in a normal conversation.

# Chapter 3

# Methodology

The design process of an application for e-textiles involves issues such as the selection of components from a vast pool of available alternatives, the placement of sensors, the interconnection between different components, and several other design variables. The worthiness of the design is measured by a design metric. The goal of the design process is to obtain the best possible design metric. Because the quality of the application depends on so many independent variables, each of the design variables has to be analyzed individually. To explore the complete design space, a combination of simulation and prototypic experiments is necessary.

The first section of this chapter details the general issues involved in the design of a new application for e-textiles. After a brief overview of the applications, the later half of the chapter then describes the modules that are involved in the design process of two applications. The first application, the large-scale beam former, estimates the direction of arrival (DOA) of a moving vehicle. The design metrics for this application are the accuracy of the estimated direction of arrival, the accuracy of the estimated coordinates of the source, and the power consumption. The second application is a speech processing vest. This wearable fabric

looks at the advantages of placing microphone arrays on different locations on the body in estimating the time difference of arrival (TDOA) for speech signals and improving the quality of the speech in a teleconferencing room. The design metric for this application is the accuracy of the estimated direction of arrival of an active speaker and the quality of the speech obtained by selectively combining signals from a subset of microphones.

## 3.1 What is Needed ?

Active research has brought to light the principle issues that are to be considered when designing an e-textile application. The field of e-textiles is itself a union of two fully developed yet totally independent technologies. This gives rise to a completely new set of problems that had not been addressed previously. Some of them are related to the physical environment for which the application is designed whereas other limitations are inherent in the textiles and electronics industries. A major challenge in the design of an e-textile is that the range of design issues spans diverse areas. Some of these are described as follows.

**Placement of sensors:** In some applications, e.g. where motion detecting sensors or microphones are involved, placement of the sensors can make a huge difference in the quality of data acquired. The quality of the application often depends upon the quality of the obtained sensor data. For example in wearable fabrics, twisting and bending of the torso can change the relative positions of the sensors. For applications involving acoustic signals, the frequency range of interest puts a limit on the distance between sensors.

**Form factor:** A wearable fabric is often made with a textile that is flexible enough to give it a comfortable form factor. When some of the threads in the fabric are replaced by wires, the rigidity of the wires can change its form factor and can make it uncomfortable

for the wearer. This is an important factor that has to be considered in the design.

**Draping of clothing:** In most of the electronic applications, the placements of the sensors are fixed at precise locations. But when these sensors are integrated into the fabric, the flexibility of the fabric changes the relative positions between these sensors. This flexibility can thus become a limiting or an advantageous factor in the applications where the placement of sensors plays an important role.

**Interconnection:** This might be one of the most difficult areas to be dealt with. In the electronics industry, an interconnection involves either connecting two wires or connecting an electronic component to a wire. The common method that is used for both these interconnections is soldering. Components can also be connected to the wires using insulation displacement connectors and spot welding. On the other hand, stitching interconnects two pieces of fabrics. When two pieces of e-textiles have to be interconnected, both these issues have to be considered simultaneously. Thus, there is a need to develop new types of interconnection between electronic components as well between the textiles. If this e-textile is a wearable fabric then care has to be taken to make it comfortable enough for the wearer.

**Power:** Most of the e-textile applications are stand-alone applications with embedded sensors, computational entities, output functionality, and power sources. Because battery life is limited, it has to be distributed in an intelligent manner. Special modules have to be included to switch the electronics on or off as and when necessary.

**Networking and communication:** In the applications where data acquisition from many sensors is involved, issues such as addressing of the individual sensors, the layout of the data paths within the fabric, the placement of the processing units, and the routing strategies all play a significant role in the design of the fabric in terms of its power consumption.

**Software and execution:** Power consumption also depends on the software execution time and the way in which the data is retrieved from extended memory. In addition, the fact that limited computational logic (simpler DSP's cannot execute instructions in parallel) can be embedded on the fabric can increase the execution time.

**Manufacturing cost:** In the field of textiles a variety of choices for types of yarns and weaving options are readily available. At the same time in the field of electronics, countless options for components, processing entities, and system software solutions exist. A successful application should be able to be manufactured in large quantities at reasonable cost. The components that can be integrated into the e-textile during the weave time can significantly reduce the cost as compared to the options which require additional manufacturing.

It is clear from the above discussion that the possible design space for an e-textile is huge. To verify that the desired quality of the application is achieved at reasonable manufacturing cost, extensive prototyping is unavoidable. At the same time, it is impossible to create prototypes for all possible design options, as the overhead cost will be unbearably high. To explore each of these design options, simulations have to be employed. Thus a combination of prototypic and simulation experiments will have to undertaken. The novelty of this thesis is that it describes a design framework for acoustic signal processing applications for e-textiles. The design space for the two applications is explored through simulations and the manufacturability of the application is verified through limited prototyping. In general, the design process can be viewed in the form of flowchart consisting of modules as shown in Figure 3.1. It is an iterative process in which each of the modules may have to be used several times, until the desired quality is achieved at a reasonable cost. Each of these blocks are explained in detail in Sections 3.2, 3.3, 3.4, and 3.5.

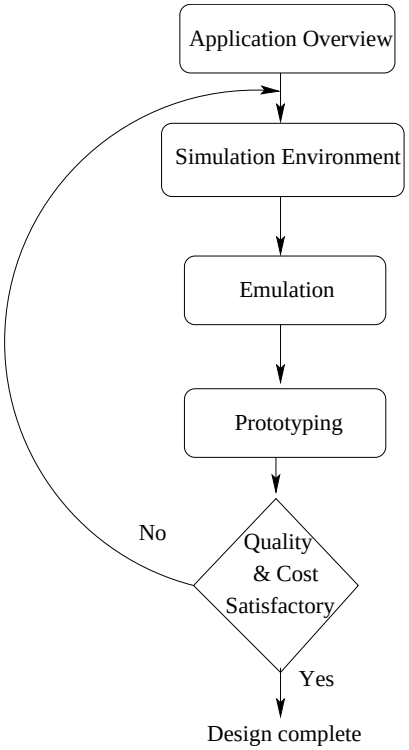


Figure 3.1: Design process

## 3.2 Applications

This thesis primarily focuses on the audio signal processing applications for e-textiles. These applications were chosen to investigate specific aspects such as placement of sensors, computational complexity of the application, power consumption, and the effect of the size of the fabric on the overall design. The two applications considered with these design variables in mind are described below.

A large-scale beam forming fabric was one of first applications that was developed using the concept of microphone arrays in e-textiles. The objective of the application was to estimate the DOA of a moving vehicle in an outdoor environment. The fabric consists of a large

number of microphones, a DSP board, and a power source. The interconnection between different components is done using flexible stainless steel wires, which were incorporated during weave time. The design process of this application investigates the accuracy of the DOA estimation of the vehicle as a function of power consumption. The design variables include separation of microphones within an array, the sampling rate, and the number of microphones per array.

In a small teleconferencing room, reverberations often degrade the acoustic signal from an active speaker. Microphone arrays can be used in such environments to improve the quality of the received speech. First an estimate of the DOA of the active speaker is obtained. The TDOA of signals received at different microphones is a required parameter in this DOA estimation. The second application considered in this thesis is a speech processing vest that tries to estimate the DOA of an active speaker inside a teleconferencing environment. The vest is designed to see the performance advantages in estimating this DOA using distributed arrays of microphones under reverberant conditions. Microphone arrays are placed on the front and the backside of this wearable vest. Because of the reverberations, the time delay estimates obtained from a pair of microphones may not be equal to the true time delay. Using the time delay estimates from more than one pair of microphones, a better estimate of the DOA can be obtained by doing a least squares fit to the TDOA's. The vest estimates the direction of arrival of the active speaker using this approach. Once the DOA for the active speaker is obtained, the received signals from a subset of microphones are combined to get a better speech quality.

## 3.3 Simulation

The goal of the simulation is to get quality values for the design metrics. The choice of design variables will be evaluated in different physical environments and in the presence of the movement of the human body and the garment. The design variables and design metrics for both applications are mentioned in Section 3.2. The simulation uses a range of values for each of these design variables and specified conditions such as the twisting of the torso, the reverberation time, and the received signal quality. The results of simulations should be able to give definite values for the design metric under the specified conditions. The values of the design variables that give the desired quality values of the design metric are used for creating prototypes.

For the applications mentioned in Section 3.2, the simulation environment consists of two sections. The first section tries to simulate the signals being sensed. For the case of acoustic applications, this involves simulating the propagation of the wave front from the acoustic source to the microphone array. Because it is a continuous signal, the underlying simulation environment should possess the notion of continuous time. These signals can then be sampled by an analog to digital conversion module and further events can be handled as discrete time events. The second section of the simulation environment does this further processing of the obtained sensor signals to give the required output. In the real application, a digital signal processor or a microcontroller does the processing.

The simulation environment should, in general, be able to model diverse physical conditions. Because the simulation has to take into account for different types sensors, as well as the nature of their responses to the varying physical conditions, the simulation tool has to allow for the development of custom models. The modeling of sensors is typically based on simple mathematical equations. In addition, the simulation tool should also be able to use the

outputs from other simulation environments, in the event that existing sensor models are already available. Popular simulation environments, like OPNET, do not have all of these features. Hence for the purpose of simulations, a heterogeneous, concurrent, and multi-domain simulation tool, Ptolemy II, is used [33]. It is currently under development at the Department of Electrical Engineering and Computer Science, University of California at Berkeley. Figure 3.2 shows a screen shot of a portion of the simulation environment.

The Ptolemy II simulation tool can be best described as follows, *"The Ptolemy project studies heterogeneous modeling, simulation and design of concurrent systems. The focus is on embedded systems, particularly those that mix technologies including, for example, analog and digital electronics, hardware and software, and electronics and mechanical devices. The focus is also on systems that are complex in the sense that they mix widely different actions, such as signal processing, feedback control, sequential decision making, and user interfaces"* [33].

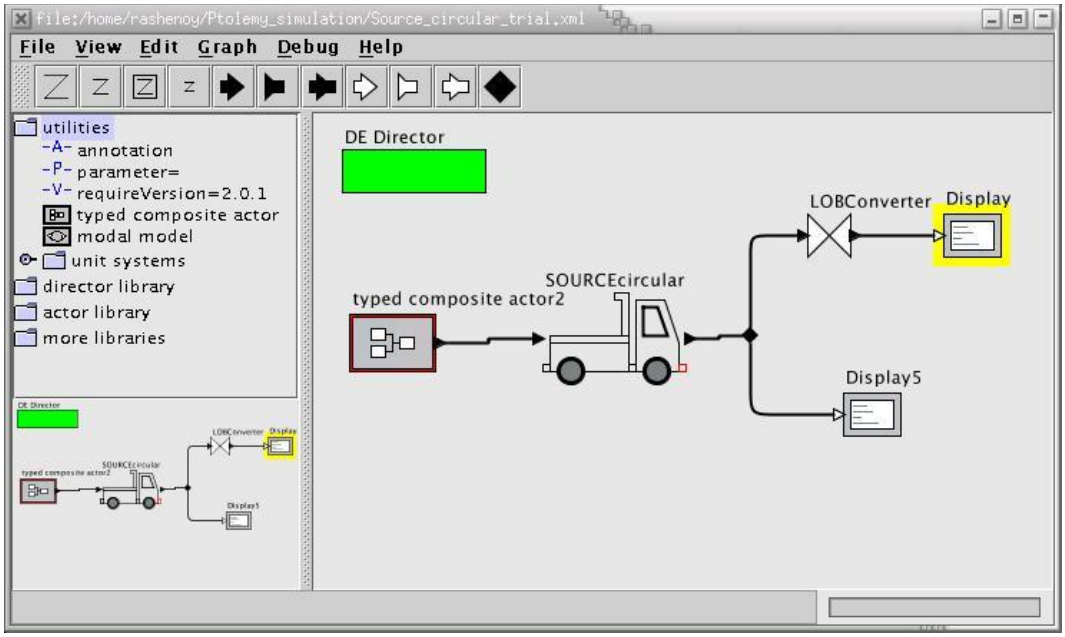


Figure 3.2: Ptolemy simulation environment

Because the simulation environment for the applications mentioned in Section 3.2 should

possess all of these features, this tool suits the type of simulations that are to be undertaken. Ptolemy II also supports construction and interoperability of components that operate simultaneously in time and have multiple sources of stimuli. Thus it can be extremely helpful in simulations where correctness of the response to stimuli and the timeliness of the response are equally important. The interactions between different components in Ptolemy follow well-defined rules. These interactions are governed by predefined semantics, known as models of computation. The most important ones are mentioned below with their brief descriptions.

**Continuous Time (CT):** This domain is specifically designed for the components that interact via continuous time signals. These components can follow specific differential or algebraic equations. It also includes a large of tools for solving differential equations. The propagation of acoustic signals from the source to the microphone array is modeled in this domain.

**Discrete Event (DE):** This model of computation governs event-based simulations. An event in this model consists of a value and a time stamp. The domain processes all the events with same time stamp using Java threads. It is particularly useful in emulating digital hardware and telecommunication systems. The emulation of a DSP or a microcontroller can be modeled in this domain.

**Finite State Machine (FSM):** The best use of this domain is in the modeling of control logic. Each component in this domain is a state. The execution is strictly based on an ordered set of state transitions. The switching devices for the applications can be modeled in this domain.

**Synchronous Data Flow (SDF):** This model is best suited for handling components that operate on streams. Most of the digital signal processing logic can be implemented in this domain.

**Discrete Time (DT):** In this domain, the communication between actors takes place with a notion of time between tokens. This model is particularly important for modeling communication actors, where the time between the tokens is uniform or where the multi-rate models have to be implemented. Networking and communication between different clusters are handled in this domain.

The most important feature of the simulation tool is its ability to embed the components of one domain into the other. Thus, it can help in modeling scenarios, in which based on specific conditions, different behaviors of the components can be activated or deactivated. Individual components can be modeled in Java using the existing Ptolemy API's. Because Ptolemy is built using JAVA, native application programs (written in C language) can be interfaced directly using Java Native Interface (JNI). This is very important in scenarios in which an existing program from a microprocessor or a DSP needs to be directly imported into Ptolemy. Ptolemy can thus address some of the issues related to the portability of programs from simulation environment to a real DSP or a microprocessor. Other features include easy I/O functionality (such as writing into a file and displaying the current status of the simulation), extensive mathematical support, extensive graphics support, support for MATLAB (interface to MATLAB engine), and output streaming of data.

## 3.4 Emulation

This step can be seen as an optional step in the design process. The cost and time required to make a complete working prototype can be very expensive. Although the values of the design variables obtained from simulation give the optimum design metric, there is a need to check the working of the entire application in more realistic conditions. This can be done by processing the signals obtained from real sensors. Actual sensors can be placed in locations

as indicated by the results of the simulations. The partial prototype can then be placed in the exact physical environment that is used for simulation. The only difference being that the signals so obtained are processed by a laptop or a desktop, instead of a DSP or microcontroller. The purpose of emulation is two-fold. As is often the case, there are a large number of DSPs and microcontrollers available in the electronics industry. The choice of the DSP or microcontroller often depends on the computational complexity of the underlying signal processing algorithm and power consumption of the processor. In addition, a lot of effort is needed to design a custom interface for data acquisition from different types of sensors. The selection of the processor can only be done once the working of the entire application is verified with actual signals. Thus, by emulating the working of a processor, this choice can be postponed to a later stage. In most of the applications, the cost of sensors is much lower than the cost of the processing entities. In addition, a data acquisition card with graphical user interface can be used to obtain data from different types of sensors. Thus, the working of the application in the real world can be verified.

## 3.5 Prototypes

This is the most important part of the overall design. The results from emulation should give a clear indication of the choice of DSP or microcontroller. Interfaces to the different sensors such as amplifiers, filters, buffers, and voltage followers have to be designed. A complete prototype will verify the successful working of the stand-alone application. It should also give an indication of the power requirements, even though these values have been obtained from simulation or emulation. Issues like the resistance of the connecting wires, the cross talk between data lines, the quality of the interconnection between the wires and sensors, and the internal noise that are not modeled through simulations can be observed through prototypic experience.

The four step iterative process needs to be repeated until the desired quality of the application is achieved at reasonable cost. The design process mentioned in this chapter was used for building the two applications. The following chapters will discuss the design process and the applications as a whole.

# Chapter 4

# Large Scale Beamformer

This chapter details the complete design process of a large-scale beamformer fabric. The first section presents a careful analysis of the application to identify all of the design variables and design metrics. The ranges of values or options for the design variables are then set based on this analysis. The next section describes the simulation environment and nature of simulations performed. The results of the simulation were used to create a prototype whose description is given in the prototype section. Finally the readings taken from a field test using an actual prototype are given in the results and conclusion section.

## 4.1 Problem statement

The goal of this application is to estimate the location of a distant moving vehicle (truck, bus, or tank) with desired accuracy using an array of microphones placed on a fabric. The microphones are to be connected to a processing unit (a DSP processor or a microcontroller) by steel threads, which will act as wires in an e-textile. Because these wires are to be incorporated during the weaving process, the placement of the microphones and the processors

have to be determined. The placement of the microphones can make a large difference in the quality of the data obtained. The quality of the data obtained can help in reducing the amount of data to be processed and can thus reduce the power consumption. Thus, the design metric for this application is the accuracy of the obtained source coordinates and the power consumption. The primary design variables are the number of microphones and the placement of the microphones.

A group of 3 to 7 microphones that are connected to a single processor is known as a cluster. The microphones are placed on the circumference of a circle, with or without a microphone at the center. Such an array is known as a circular array. The radius of the circle on which the microphones are placed is a design variable for this application. A cluster estimates the direction of arrival of a distant vehicle by using the relative time delays between different microphones. Even though the receiver does not know the exact nature of the radiated signal, the range of frequencies of the signal radiated by a bus or a tank is known *a priori* (typically 100 - 180 Hz). The spectrum of the signal recorded from a tank is shown in Figure 4.1. This corresponds to a wavelength of about 6 ft to 11 ft in the atmosphere. This restricts the maximum separation of microphones from the center of the cluster to $radius \leq 6/2$ (from Equation 2.11). Because the source is located at a distance of 50 to 200 ft, and the separation between the microphones is less than 3 ft, the far field case is assumed. As the distance between the source and the microphone array is much greater than the dimension of the array, the signal at each and every microphone is attenuated equally. Because the source is in the same plane as the sensor array and the far field case is assumed, the relative delays between the signals received at different sensors are functions of the angle of arrival only.

A two-step process used to estimate the location of a vehicle. The first step estimates the direction of arrival for a given cluster. Using the estimated direction of arrival from more than one cluster, and the coordinates of the clusters themselves, a least squares solution can

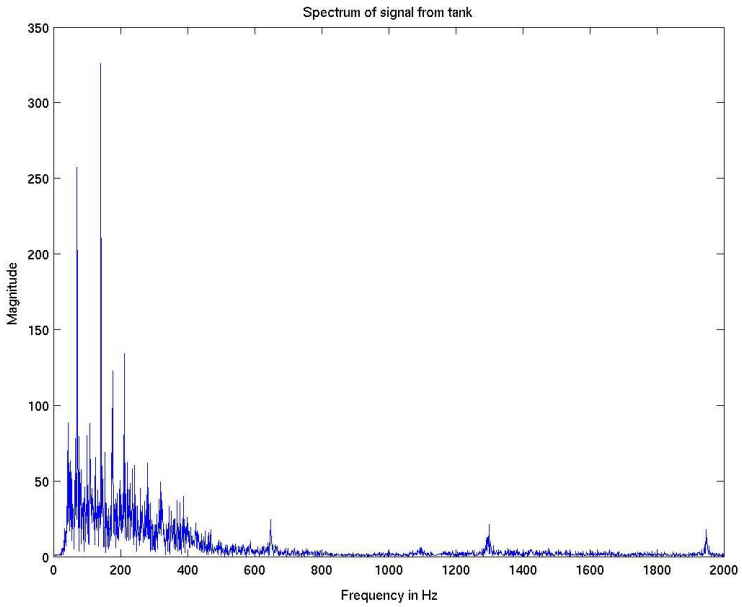


Figure 4.1: Spectrum of sound from a tank

be obtained for the location of the vehicle. This process of estimating the location of the source is done in the second step.

In the absence of the power consumption design metric, the placement of the microphones would just be governed by the width of the fabric and frequency range of interest. With no power constraints, the desired accuracy can be obtained without optimizing the placement of microphones. Because the power consumption is an equally important metric, higher accuracy should not be obtained by doing more computations and processing more data. Thus, to reduce amount of data to be processed and the number computations, the radius of the circle on which microphones are placed, the sampling rates, the distance between the clusters, and the number of clusters all become secondary design variables.

In addition to beamforming, the fabric should also have ability to switch off all of the processors temporarily and migrate to a low power state. This is very vital in scenarios where there are very few vehicles passing by. Thus, the processor that has the ability to

switch states from an active state (running all the processes) to passive or idle state is best suited for this application. Fixed point processors are very power efficient when compared to floating point processors. Therefore, the underlying algorithm should be implemented in fixed point arithmetic.

## 4.2 Simulation

The goal of simulation is to get the values of the design metrics for the specified set of design variables under various physical environments. As mentioned in the previous section, the tradeoff is between accuracy and power consumption. The simulations are done in Ptolemy II for the reasons given in Section 3.3. The placement of microphones on the circumference of the circle is shown in Figure 4.2. The simulation consists of three components that are described in the following subsections.

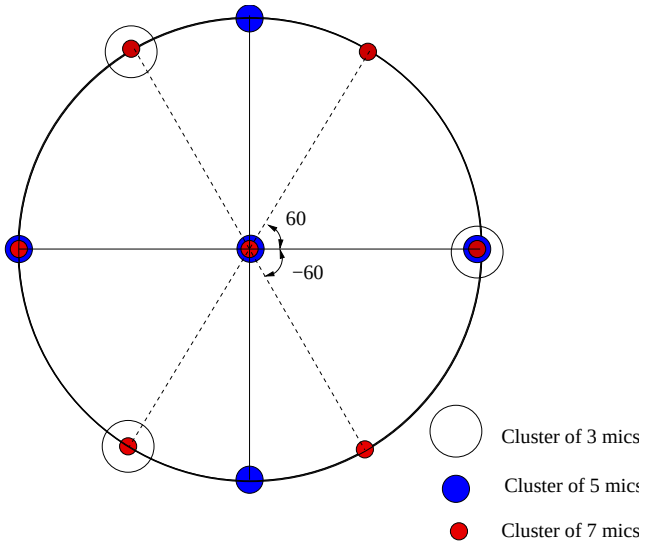


Figure 4.2: Placement of microphones

### 4.2.1 Wave Propagation

The first component simulates the actual signals seen by the microphones in an outdoor environment emanating from a single active source in the far field. An acoustic source in the far field can be stationary or moving at a constant speed. The source is also assumed to be moving either along a straight line or along the circumference of a circle. The circular trajectory is particularly important in characterizing the accuracy of the direction of arrival estimation. The motion of the vehicle along the straight line is a more realistic scenario. The signals seen by the microphones are attenuated and phase delayed as mentioned in Section 2.4. This step involves the calculation of the path lengths from the source to each of the microphones and then computing the actual phase delays based on the frequency of the signal. Even though attenuation is a function of the actual path length, the signals strengths at all of the microphones are equal because of the far field assumption. These signals are simulated in the continuous domain.

### 4.2.2 A/D Conversion and DOA Estimation

This component consists of an analog-to-digital (A/D) conversion module followed by a direction of arrival estimator. An A/D conversion module samples the signal generated by the first component. These sampled values are then quantized with 8-bit precision. As mentioned Section 4.1, the sampling rate of the A/D is a design variable and can take values from 1024 to 8192 samples per second. These quantized samples are then used to estimate the direction of arrival of the active source. The direction of arrival estimation is done using a simple delay-and-sum beamformer described in Section 2.6. The beamforming algorithm is derived from an Army Research Laboratory algorithm modified to achieve low power consumption [34]. This beamformer is computationally very inexpensive as there are

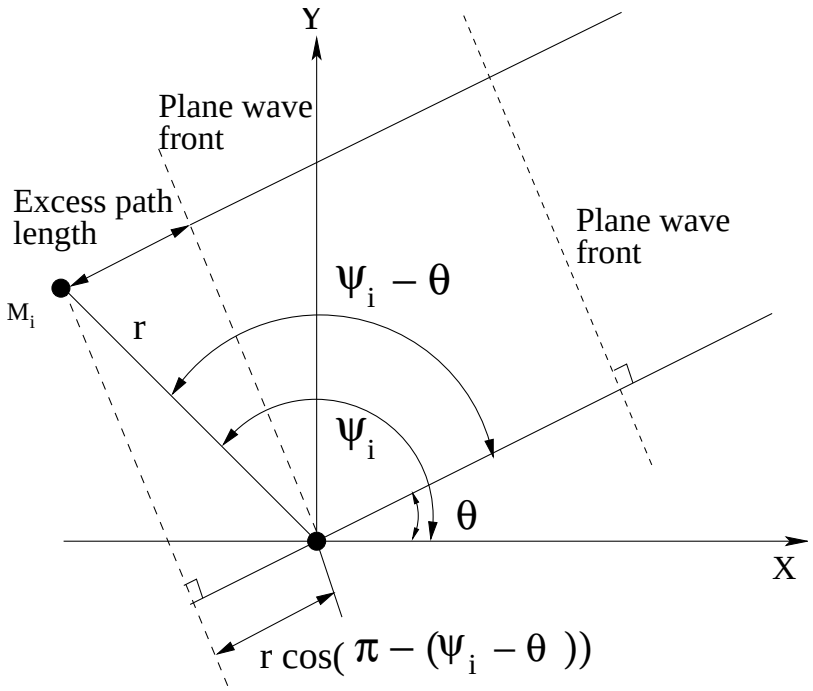


Figure 4.3: Time difference of arrival for circular array

no complex multiplications or matrix manipulations involved. The delays for different angles of incidence can be pre-calculated and stored in memory. With the center of the circle as the origin, the polar coordinates of the $i^{th}$ microphone are given by $\{r, \psi_i\}$. Because a circular array is used for this application, r is a constant for all of the microphones except for the one at the center. From Figure 4.3, it can be seen that the actual delay for the $i^{th}$ microphone with respect to center of the circle for an angle of incidence $\theta$ is given by

$$\begin{aligned} \tau_i &= \frac{r\ cos(\pi - (\psi_i - \theta))}{c} \\ &= \frac{r\ cos(\theta - \psi_i)}{c}. \end{aligned} \tag{4.1}$$

These actual delays are then expressed in an integer number of samples. As the delays are just functions of the angles of arrival, once these values are calculated at the beginning of the simulation they can be reused. Once the necessary samples are obtained from the A/D conversion module, the obtained discrete time signals are delayed by the corresponding

sample delays for different angles of incidence. Changing the indices within the array of samples incorporates the required delay in the time domain. These delayed versions of the signals are then added. The entire procedure is shown in Figure 4.4.

The search for peak power is done for all integral values of incident angles of $\theta$ i.e. $0 \leq \theta \leq 359$. This method of peak power search is computationally expensive as the number of search angles is quite large. This search can be reduced significantly by finding the power only at increments of five degrees i.e. 0, 5, 10 ... and then using a spline function to get the intermediate values. As the data samples for the microphones are fixed-point numbers, the estimation of DOA is done using fixed-point arithmetic. Although delay-and-sum may not be the most accurate beamforming method, the computational complexity of this method is substantially less than any of the other beamforming methods. This is the primary reason for selecting this method for the application.

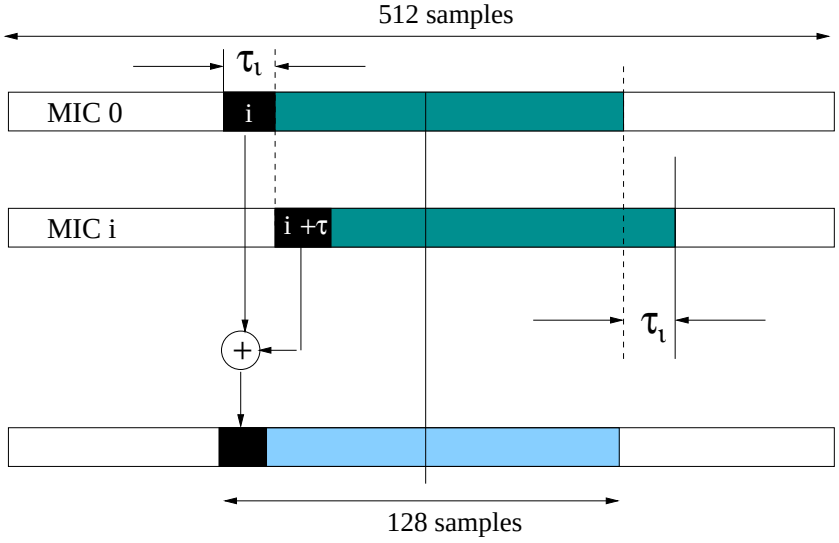


Figure 4.4: Implementation of delay-andsum beamformer

### 4.2.3 Triangulation

This component covers the least squares method to localize the vehicle using the DOA estimates from more than one cluster. Each cluster of 3, 5, or 7 microphones will give the DOA estimate. In order to localize a source, it is assumed that the underlying hardware has the communication protocol to exchange the DOA estimates. All of the clusters also know the location of other clusters *a priori*. Even if the exact location is not known, an approximate value of the separation of the clusters can be used to get the source coordinates. For the purpose of simulations 2 to 4 clusters are used at different separations ranging from 8 ft to 16 ft. All of the clusters are assumed to be in a straight line. The least squares fit for the location estimate can be formulated as follows. As shown in Figure 4.5, N clusters are being used to estimate the coordinates of the source. $\{x_i, y_i\}$ denote the coordinates of the

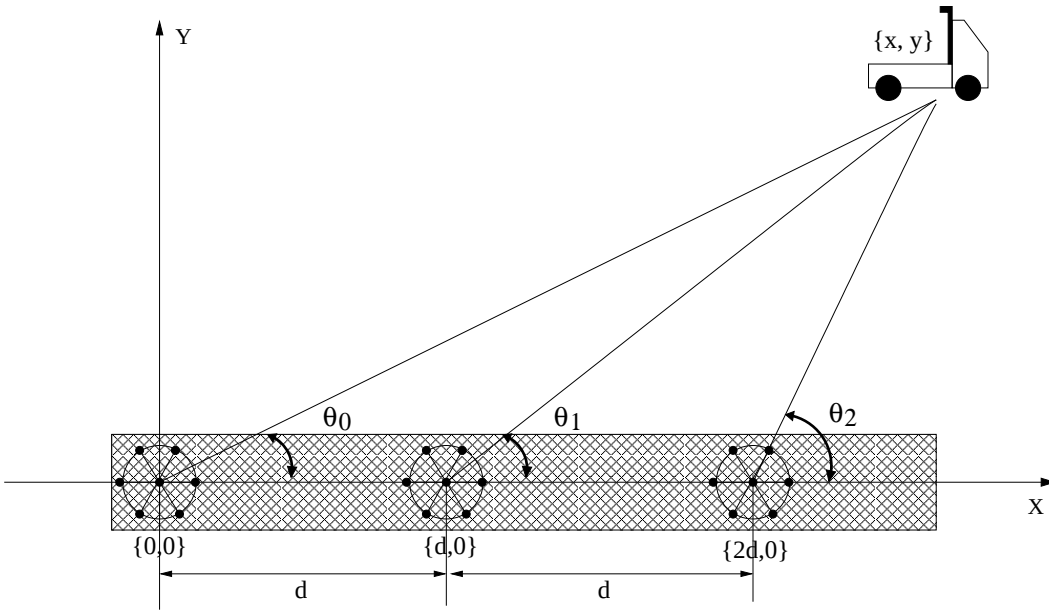


Figure 4.5: Source localization using array of clusters

$i^{th}$ cluster. Each cluster will give a DOA estimate $\theta_i$ as shown. The equation for the line of

bearing for the $i^{th}$ cluster can be written as

$$\begin{aligned} tan(\theta_i) &= \frac{y - y_i}{x - x_i} \qquad \text{and} \\ y - tan(\theta_i)x &= y_i - tan(\theta_i)x_i. \end{aligned} \tag{4.2}$$

The equations for line of bearing for other clusters can be written in a similar way. Thus, when more than two clusters are available, a least squares solution can be obtained for the source coordinates as follows

$$\begin{aligned} \begin{bmatrix} y - tan(\theta_1)x \\ y - tan(\theta_2)x \\ \dots \\ y - tan(\theta_N)x \end{bmatrix}_{N\text{x}1} &= \begin{bmatrix} y_1 - tan(\theta_1)x_1 \\ y_2 - tan(\theta_2)x_2 \\ \dots \\ y_N - tan(\theta_N)x_N \end{bmatrix}_{N\text{x}1} \\ \begin{bmatrix} 1 & -tan(\theta_1) \\ 1 & -tan(\theta_2) \\ \dots & \dots \\ 1 & -tan(\theta_N) \end{bmatrix}_{N\text{x}2} \begin{bmatrix} y \\ x \end{bmatrix}_{2\text{x}1} &= \begin{bmatrix} y_1 - tan(\theta_1)x_1 \\ y_2 - tan(\theta_2)x_2 \\ \dots \\ y_N - tan(\theta_N)x_N \end{bmatrix}_{N\text{x}1} \end{aligned} \tag{4.3}$$

$$\mathbf{AX} = \mathbf{B} \tag{4.4}$$

$$\mathbf{B} = \begin{bmatrix} y_1 - tan(\theta_1)x_1 \\ y_2 - tan(\theta_2)x_2 \\ \dots \\ y_N - tan(\theta_N)x_N \end{bmatrix}_{N\text{x}1} \quad \mathbf{A} = \begin{bmatrix} 1 & -tan(\theta_1) \\ 1 & -tan(\theta_2) \\ \dots & \dots \\ 1 & -tan(\theta_N) \end{bmatrix}_{N\text{x}2} \quad \mathbf{X} = \begin{bmatrix} y \\ x \end{bmatrix}_{2\text{x}1} \tag{4.5}$$

$\mathbf{X}$ is the unknown which can can be found by solving

$$\left[\mathbf{A}^T\mathbf{A}\right]_{2\text{x}2} \mathbf{X} = \mathbf{A}^T\mathbf{B}. \tag{4.6}$$

Assuming $\left[\mathbf{A}^T\mathbf{A}\right]_{2\text{x}2}$ is non-singulare a least square solution for $\mathbf{X}$ can be found. Values for the number of clusters and the separation between the clusters that give the desired accuracy for the location estimate have to be found through simulation.

## 4.3 Results

This section presents the results of the simulation for the accuracy of the DOA estimate and the triangulation algorithm.

### 4.3.1 Simulation Results for the DOA Estimation

The accuracy of the estimated DOA depends upon four variables: sampling rate, placement of microphones, number of microphones per cluster, and the signal-to-noise ratio of the signal received. All of these factors directly or indirectly affect the power consumption of the fabric. The quality of the DOA estimated as a function of each of these four design variables is presented in the following section. The last subsection discusses the power consumption of the fabric.

#### Simulation scenario

In order to determine the accuracy of the estimated DOA, an acoustic source is positioned in the far field at a fixed distance of 50 ft at different angles of incidence. The acoustic source in this case is a battlefield tank. As discussed in Section 4.1, energy density of the signal emitted by a tank is highest in the range of 100 Hz to 180 Hz. The acoustic source in the simulation is emitting a superposition of three frequencies at random within 100 Hz to 180 Hz. The amplitude and the initial phase of each of these sinusoids are also picked at random at runtime. As mentioned in Section 4.2.1, the phases seen by each of the microphones are calculated at each and every sampling instant. These phase-delayed sinusoids are then superimposed and white Gaussian noise is added to the sum of sinusoids. This signal is then sampled by the A/D converter and passed on to the DSP for further processing.

**Results**

The error in the estimated DOA, $\epsilon$, is the difference between the actual angle of incidence and the estimated DOA. The accuracy of the estimated DOA is reported in terms of the mean of the absolute error $E\{|\epsilon|\}$, the standard deviation of the error $E\{|\epsilon|^2\} - E\{|\epsilon|\}^2$, and the maximum error $max\{|\epsilon|\}$.

**Effect of sampling rates in ideal case:** The possible values of the sampling rates are 2048, 4096, and 8192 samples per second. The circumference of the circle on which microphones are placed is initially fixed to the maximum possible value of 3 ft. DOA estimation using 3, 5, and 7 microphones for each of the three possible sampling rates is simulated. Signals coming from the acoustic source are assumed to be free of any

Table 4.1: Accuracy of DOA estimation with radius = 3 feet in noiseless case

| Number of Microphones | Sampling Rate | Absolute Average | Standard Deviation | Maximum Error |
|---|---|---|---|---|
| **3** | 8192 | 0.2083 | 0.467 | 1 |
| | 4096 | 0.3 | 0.4552 | 1 |
| | 2048 | 0.6471 | 0.8989 | 3 |
| **5** | 8192 | 0.2222 | 0.4163 | 1 |
| | 4096 | 0.3 | 0.4589 | 1 |
| | 2048 | 1.133 | 1.4072 | 4 |
| **7** | 8192 | 0.175 | 0.3805 | 1 |
| | 4096 | 0.325 | 0.469 | 1 |
| | 2048 | 0.6833 | 0.9195 | 3 |

noise and degradation due to multipaths. The results of the simulation are shown in Table 4.1. From the simulation results, it can be seen that the accuracy of the DOA estimate increases as the sampling rate is increased. In this ideal noise free case, an increase in the number of microphones does not improve the accuracy of the DOA estimate by a significant amount.

**Degradation of accuracy with noise:** When additive white Guassian noise is added to the sum of the sinusoids, the signal-to-noise ratio of the received signal decreases. To see the effects of noise, a second simulation with a signal-to-noise ratio of 5 dB is carried out. From the simulation results in Table 4.2, it can be seen that the accuracy

Table 4.2: Accuracy of DOA estimation with radius = 3 feet with 5dB SNR

| **Sampling Rate** | **Number of Microphones** | **Absolute Average** | **Standard Deviation** | **Maximum Error** |
|---|---|---|---|---|
| **8192** | 3 | 1.8639 | 2.2807 | 7 |
| | 5 | 1.5722 | 1.9443 | 6 |
| | 7 | 1.047 | 1.353 | 5 |
| **4096** | 3 | 2.35167 | 2.8667 | 8 |
| | 5 | 1.8472 | 2.4641 | 8 |
| | 7 | 1.4611 | 1.8 | 5 |
| **2048** | 3 | 2.4194 | 3.05 | 10 |
| | 5 | 2.2889 | 2.85 | 8 |
| | 7 | 1.5889 | 1.99 | 7 |

of the DOA estimate for all of the clusters has decreased because of the presence of additive Gaussian noise. As compared to the noise-free case, it can also be seen that

the accuracy of the 3-microphone and the 5-microphone clusters has degraded more drastically than the 7-microphone cluster.

**Reduction in radius:** Because one goal of the application is to make the cluster small, the radius of the cluster is decreased to 1 ft. As said earlier, in the noise-free case, an increase in the number of microphones does not increase the accuracy of the DOA estimate. Therefore, noise is added in this simulation. The performance of all of the three clusters with different sampling rates is shown in the Table 4.3.

Table 4.3: Accuracy of DOA estimation with radius = 1 feet with 10 dB SNR

| **Sampling Rate** | **Number of Microphones** | **Absolute Average** | **Standard Deviation** | **Maximum Error** |
|---|---|---|---|---|
| **8192** | 3 | 2.9 | 3.727 | 11 |
| | 5 | 2.85 | 3.49 | 11 |
| | 7 | 1.78 | 2.189 | 6 |
| **4096** | 3 | 3.7333 | 4.5967 | 15 |
| | 5 | 3.8 | 4.53 | 12 |
| | 7 | 2.88 | 3.542 | 10 |
| **2048** | 3 | 7 | 8.32 | 18 |
| | 5 | 6.4 | 7.7 | 17 |
| | 7 | 5.6 | 6.7 | 13 |

The decrease in distance reduces the accuracy. In a delay-and-sum beamformer the signals are shifted by integer numbers of sample delays, where each sample delay represents the time between two sampling instants in real time. The time shifts resulting

from integral sample delays can have values given by

$\frac{k}{Fs}$ seconds where Fs is the sampling rate and k = 0, 1, 2 ...

If the actual delay is an intermediate value such that

$$\frac{k-1}{Fs} < \frac{d_{ij}cos(\theta)}{c} < \frac{k}{Fs},$$

for any integral value of $k$, then that delay is rounded off to the nearest integer sample delay. For a fixed radius, the maximum value of the delay can be $\pm d/c$. When the radius is decreased from 3 ft to 1 ft, the maximum delay also decreases by the same factor. This results in fewer possible integer values which $k$ can take. Each unique DOA estimated will have a unique combination of the delays for all of the microphones. Thus, when the radius is decreased, the number of unique combinations for delays also decreases and hence the decrease in accuracy.

For example, consider the initial 3 ft case with 2048 samples per second. The maximum delay between the signals arriving at any of the microphones on the circumference and that of the center of the cluster can be

$$\frac{\text{radius of the cluster}}{c} = \frac{3ft}{1090ft/sec} = 2.75ms.$$

This when converted into sample delays, results in a maximum value of

$$max\ int\ \leq \{2.75ms * 2048\} = 5$$

integer sample delays; the range is $\{-5, -4, \ldots, 4, 5\}$. When the distance is reduced to 1 ft, the possible sample delays will have to be less than

$$max\ int\ \leq \{\frac{2048 * 1ft}{1090ft/sec}\} = 2$$

with the range being $\{-2, -1, 0, 1, 2\}$. Thus, because of the reduction in the sample delays that can be used, the possible combinations of sample delays for different microphones have decreased. As a result, the numbers of distinct DOA's that can be

estimated have decreased. Thus, the mean error is more in the 1 ft case. A plot showing the estimated DOA's for angles of incidence between 45 degrees and 135 degrees for the case of 3 ft and 1 ft is shown in Figure 4.6. For the same reason the higher

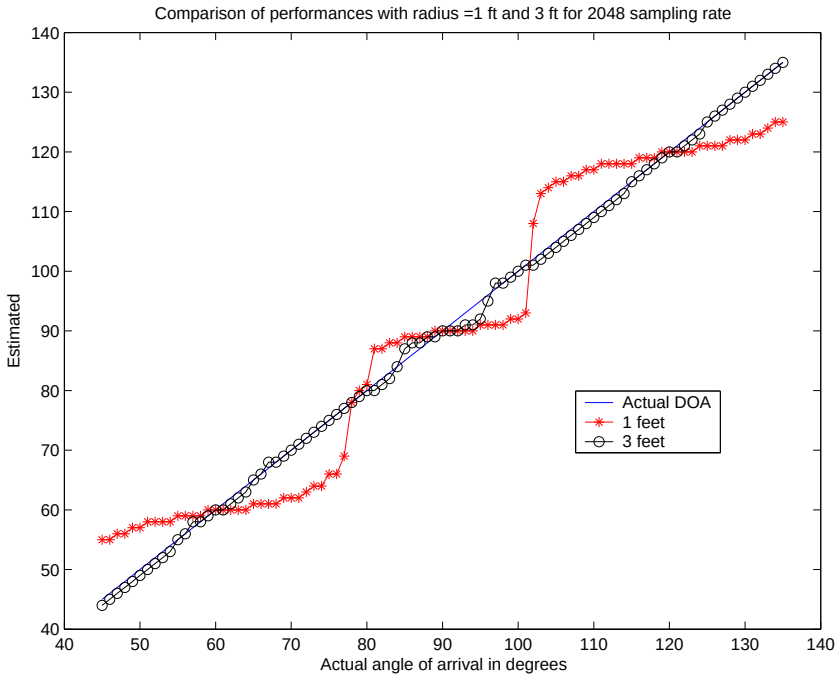


Figure 4.6: Comparison of estimated DOA's for 3 ft and 1 ft cluster with 2048 sampling rate

sampling rates give better DOA estimates. In presence of noise, as we increase the sampling rate the performance of 7-microphone improves by a greater amount as compared to 3 and 5 microphone case. Therefore, the 7-microphone cluster is the preferred choice for developing prototypes.

**Signal-to-Noise Ratio:** Finally, the performance of the 7-microphone cluster is tabulated as a function of the sampling frequency and SNR. Clearly, the DOA estimate improves with better SNR and higher sampling rates.

Table 4.4: Accuracy of DOA estimation as a function number of SNR with radius = 1 ft

| **Sampling Rate** | **SNR (dB) dB** | **Absolute Average** | **Standard Deviation** | **Maximum Error** |
|---|---|---|---|---|
| **8192** | 7 | 2.3 | 2.9277 | 11 |
| | 10 | 1.78 | 2.189 | 6 |
| | 13 | 1.0472 | 1.3423 | 5 |
| | 19 | 0.5417 | 0.6057 | 3 |
| | no noise | 0.275 | 0.447 | 1 |
| **4096** | 7 | 3.9972 | 4.988 | 16 |
| | 10 | 2.88 | 3.543 | 13 |
| | 13 | 2.25 | 2.7794 | 10 |
| | 19 | 1.5889 | 1.9117 | 5 |
| | no noise | 1.1861 | 1.4196 | 4 |
| **2048** | 7 | 7.3278 | 8.7603 | 23 |
| | 10 | 5.6 | 6.7 | 16 |
| | 13 | 6.0472 | 7.0661 | 16 |
| | 19 | 5.5639 | 6.463 | 13 |
| | no noise | 5.13189 | 5.9613 | 11 |

**Power consumption**

An increase in the accuracy of the DOA estimate with an increase in the sampling rate is accompanied by an increase in the power consumption. As explained in Section 4.2.2, the delay-and-sum beamformer adds the delayed signals from all of the microphones. This summation is done over a finite period. Because the frequency range of interest is $\geq 100Hz$, the minimum time over which the signals from all of the microphones should be added should be greater than the period of this wave i.e. $> 0.01s$. For our purposes, the delay-and-sum beamformer sums the signals over three periods of the wave. This corresponds to $3*0.01/Fs$ number of samples. Hence, 64, 128, and 256 samples have to processed for 2048, 4096, and 8192 Hz sampling frequencies respectively. Thus, an increase in sampling rate results in an increase in the number of data points to be processed and correspondingly an increase in the power consumption. Typical A/D conversion units consume more power for higher sampling rates. Both of these factors directly result in an increase in the power consumption.

### 4.3.2 Simulation Results for Triangulation

The ultimate goal of the application is to estimate the x and y coordinates of the moving vehicle. Using the DOA estimates from more than one cluster, the location of the vehicle can be estimated by a least squares fit as explained in Section 4.2.3. The accuracy of the location estimate depends upon the accuracy of the DOA estimate and the separation between the clusters.

**Simulation Scenario**

For the purpose of this simulation, a tank is assumed to be moving in a straight line with a constant velocity. As shown in Figure 4.7, a fabric with 2, 3, and 4 clusters is kept at a distance of 50 ft from the moving vehicle. Each cluster consists of 7 microphones. The design variables that we are interested in include the number of clusters on the fabric and the separation between the clusters. For getting good location estimates both the distance

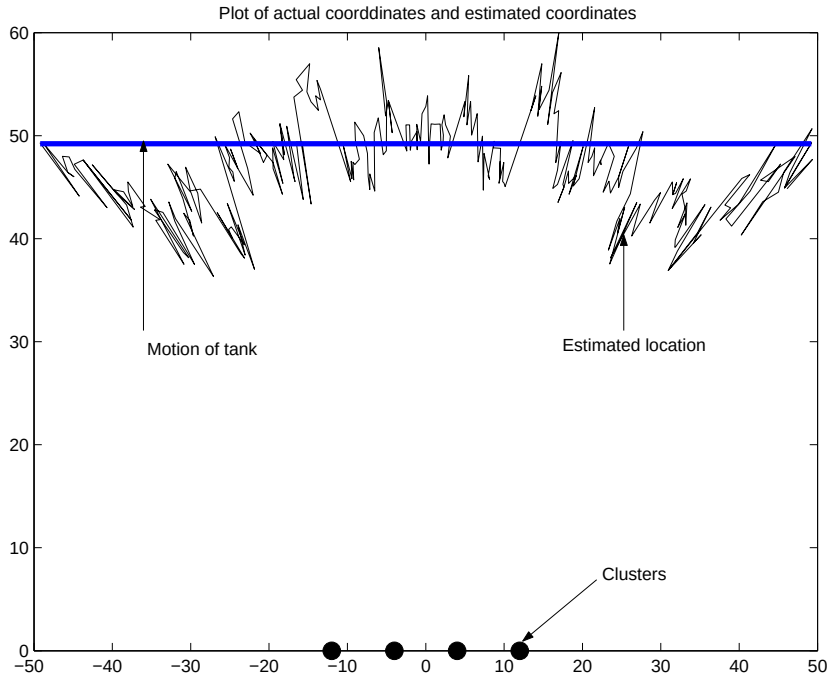


Figure 4.7: Estimation of the location using 4 cluster at {-12,-4,4,12} with 8192 sampling rate

and the angle have to be accurate. A small derivation shows how small changes in the DOA estimates from one of the clusters affect the distance estimate. As shown in Figure 4.8, $\theta$ is the angle of incidence on the cluster at the origin and $\alpha$ is the DOA estimate from cluster

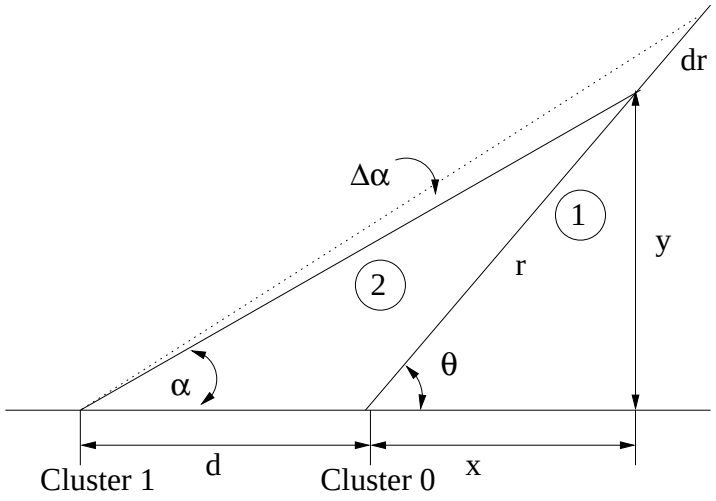


Figure 4.8: Error in location estimate with respect to error in DOA estimate

at $\{-d, 0\}$. The equations of line 1 and line 2 are

$$y = x\ \tan(\theta), \qquad \text{and} \tag{4.7}$$

$$y = x\ \tan(\alpha) + d\ \tan(\alpha). \tag{4.8}$$

Solving for x and y we get

$$\begin{aligned} y(\tan(\theta) - \tan(\alpha)) &= d\ \tan(\theta)\tan(\alpha), \\ y &= \frac{d\ \tan(\theta)\tan(\alpha)}{\tan(\theta) - \tan(\alpha)}, \qquad \text{and} \end{aligned} \tag{4.9}$$

$$x = \frac{d\ \tan(\alpha)}{\tan(\theta) - \tan(\alpha)}. \tag{4.10}$$

$r$ is the estimated distance of the vehicle from the origin when $\alpha$ does not have any error.

$$\begin{aligned} r &= \sqrt{(x^2 + y^2)} \\ &= \frac{d\ \tan(\alpha)\sec(\theta)}{\tan(\theta) - \tan(\alpha)} \end{aligned} \tag{4.11}$$

A small change in alpha, $\Delta\alpha$, can change the estimated location. Differentiating $r$ with respect to $\alpha$ will give the change in $r$ when $\alpha$ is changed by $\Delta\alpha$

$$\frac{\partial\, r}{\partial\, \alpha} = \frac{d\, \sin(\theta) \sec(\alpha)^2}{(\tan(\theta) - \tan(\alpha))^2}. \tag{4.12}$$

Thus, it can be seen that the error in the estimated distance is very large when the estimated DOA of cluster 1 is very close to the actual angle of incidence at the origin. The DOA estimate, $\alpha$, will be very close to $\theta$ when the separation between the two clusters is small. In the other words, to improve the location estimates, the separation between the clusters should be large and the DOA estimates should have as little error as possible. This is true for any number of clusters. When more than two clusters are used, the DOA estimates of all the clusters have to be good; otherwise, the best fit estimated coordinates will also have large error.

The other factor which should be taken into account is that, when the actual angle of incidence $\theta$ is small, the error in estimated distance is large. This is because when $\theta$ is small, $\tan(\theta)$ and $\tan(\alpha)$ are very small and the difference between them will be still smaller. For the same reason, a fabric with a linear array of clusters can give good location estimates when the angle of incidence is closer to $\pi/2$. It is for this reason that, in the simulation, the start and the end position of the tank will have angles of incidence between $\pi/4 \leq \theta \leq 3\pi/4$.

For angles of incidence within $\pi/4$ and $3\pi/4$ and when the accuracy of the DOA estimate is good, the error in estimated distance for smaller cluster separation is as good as those for larger cluster separation. This is because, $\partial r/\partial\alpha$ from Equation (4.12) is directly proportional to the separation between the clusters. Thus when DOA estimates are good, the error in distance is more for larger separation between the clusters.

In order to quantify the quality of the results of the location estimate, a single measure of accuracy is required. But a true picture of location estimates can be obtained only when both the distance and the DOA are taken into account. Therefore, using the estimated x

and y coordinates, an estimated distance and angle in polar coordinates is calculated. This angle is called $\text{DOA}_{LS}$. The ratio of the difference between estimated distance and the actual distance to the actual distance, the mean of the absolute value, the standard deviation, and the maximum error between the actual angle of incidence and $\text{DOA}_{LS}$ is reported.

**Result**

The results of the simulation to observe the effects of the number of clusters and the separation between the clusters are now presented.

**Separation between the clusters:** From the discussion in Section 4.3.2, it is evident that the placement of the clusters plays an equally important role as that of the DOA estimates. The first set of simulations were done with two clusters at x coordinates {-8,8}, {-12,12}, and {-16, 16} ft. The velocity of the vehicle is 5 m/sec and is moving parallel to the x-axis at a distance of 50 ft. The results of the simulation are shown in Table 4.5.

It can be seen that when the separation between the clusters is small, the distance and the $\text{DOA}_{LS}$ estimates are as good as those for the larger separation. From Section 4.3.1, we know that the DOA estimates improve with an increase in the sampling rates. Thus, for the higher sampling rates the location estimates for all of the three cases are equally good. When the sampling rate is 4096 Hz, the DOA estimates are very good. This results in better distance estimate for lower separation as compared to higher separation between the clusters. When the sampling rate is reduced to 2048 samples per second, the clusters at {-8, 8} are unable to get the location estimates as the estimated DOA's of both the clusters are almost identical. In this case, wider separation is more useful.

**Number of clusters:** To see the effect of more clusters on the location estimates, three

Table 4.5: Accuracy of location estimation using 2 clusters

| **x-coordinate of clusters** | **Sampling Rate** | **Error Distance** | **DOA$_{LS}$** | | |
|---|---|---|---|---|---|
| | | | **Absolute Average** | **Standard Deviation** | **Maximum Error** |
| **[-8, 8]** | 8192 | 0.6813 | 0.995 | 1.145 | 3.22 |
| | 4096 | 0.63 | 1.73 | 2.1434 | 4.6 |
| | 2048 | - | - | - | - |
| **[-12, 12]** | 8192 | 0.6716 | 0.9667 | 1.084 | 2.606 |
| | 4096 | 0.66559 | 2.46 | 2.73 | 5.308 |
| | 2048 | 5.3375 | 5.9192 | 7.362 | 14.6602 |
| **[-16,16]** | 8192 | 0.67272 | 0.87589 | 1.012 | 2.642 |
| | 4096 | 0.87 | 2.49 | 3.30329 | 6.2157 |
| | 2048 | 4.002 | 5.206 | 6.4961 | 12.77 |

and four clusters are used in this simulation. The x-coordinates of the three and four clusters are {-12,0,12} and {-12,-4,4,12}. The speed of the vehicle and the location of the vehicle is same as in the previous case. In this case, the location estimates using three and four clusters are better than the using two clusters considered in the previous section. This is true only for higher sampling rates when the DOA estimates of all the clusters are good. Under low sampling rates the distance estimates tend to be better but the DOA$_{LS}$ is worse than using 2-two clusters as the error in DOA estimates of all the clusters are poor. The performance of three clusters is worse than the two cluster case but better than four cluster case. This is true for low sampling rates and the behavior is exactly opposite in the higher sampling rate case.

A plot representing the DOA estimates for various angles of incidence is shown in

Table 4.6: Comparison of location estimation using 2, 3 and 4 clusters

| **x-coordinate of clusters** | **Sampling Rate** | **Error Distance** | **$DOA_{LS}$** | | |
|---|---|---|---|---|---|
| | | | **Absolute Average** | **Standard Deviation** | **Maximum Error** |
| **[-12,-4,4,12]** | 8192 | 0.65662 | 0.761 | 0.81988 | 1.8417 |
| | 4096 | 0.66847 | 1.064 | 1.1862 | 2.8709 |
| | 2048 | 4.088 | 7 | 8.157 | 14.461 |
| **[-12,0,12]** | 8192 | 0.657 | 0.843 | 0.959 | 2.3331 |
| | 4096 | 0.65822 | 1.5233 | 1.79 | 3.24 |
| | 2048 | 4.3917 | 6.63 | 7.839 | 14.5 |
| **[-12,12]** | 8192 | 0.6716 | 0.9667 | 1.084 | 2.606 |
| | 4096 | 0.63 | 2.46 | 2.73 | 5.308 |
| | 2048 | 5.3375 | 5.9192 | 7.362 | 14.6602 |

Figure 4.9. The actual DOA for the moving vehicle is the continuous line in the center. The $DOA_{LS}$ estimate using two clusters at $\{-12, 12\}$ follows the actual DOA more closer than $DOA_{LS}$ case when four clusters are used. This is only true for a sampling rate of 2048 Hz.

## 4.4 Description of the Prototype

The VT e-textiles group, with support from DARPA, has constructed an acoustic beamforming prototype textile that allows us to measure the accuracy of the DOA and the location estimate and power of each subsystem [35]. The prototype is based around a DSP-based

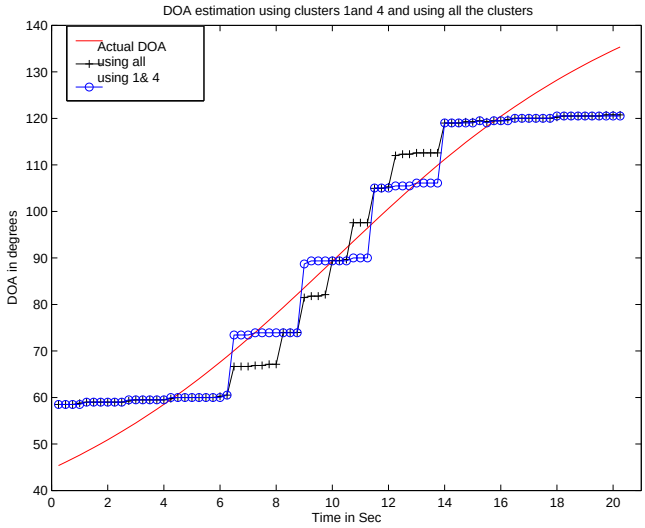


Figure 4.9: Comparison of estimated DOA's using clusters 1 and 4 and using all 4 clusters with 2048 sampling rate

data collection/processing unit that attaches to the textile. The data collection unit for each cluster is capable of sampling up to seven microphones and communicating with up to six other such data collection/processing units. The primary components of the unit are an Analog Devices ADSP-2188M, the AD7888 A/D converter, and analog circuits for interfacing to each of the microphones. The operating voltage for the components of this board is 2.75V. A low-pass filter and op-amp circuit is applied to each microphone input that is then passed to the A/D converter. The acoustic beamformer collects data when one or more microphones are detecting signals of a sufficiently high level. The entire system actively collects and analyzes the data. In addition to the data collection/processing unit, the textiles have multiple microphones and batteries attached to wires woven alongside cotton fibers. Six microphones are placed along the circumference of a circle of radius 1 ft and one microphone is at the center of the circle. The wires are strong, flexible structures constructed from many 100 % stainless steel fibers with a resistivity of 10 Ohms per meter. A photograph of a single cluster textile is given in Figure 4.10. This prototype allows for the power characteristics

of the system in operation to be determined, including the cost of analog interface circuits, data collection, data processing, and communication between units. The data processing

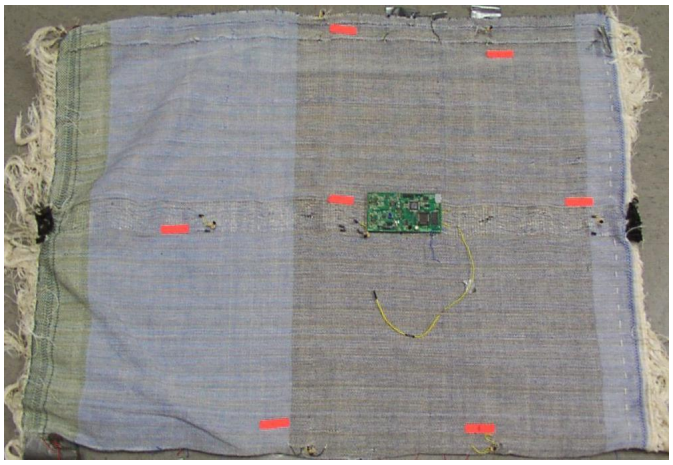

Figure 4.10: Photograph of a single cluster e-textile beamformer

consists primarily of running a beamforming algorithm on the given data samples. The beamforming algorithm estimates the direction of an acoustic source given a set of samples from the 7 microphones. The beamforming algorithm that we have chosen is derived from an Army Research Laboratory algorithm modified to achieve low power consumption [34]. This modified algorithm is described in Section 4.2.2. The beamforming algorithm is run on the most recently collected set of samples from the microphones; typical sample lengths are 512, 1024, and 2048 per microphone. The large scale beamformer has four such clusters placed on a long fabric. The clusters are separated by 8 ft. A single data bus, made up of steel wires is used for communication between these clusters. The location of the moving vehicle is estimated on any one of the clusters, after getting the DOA estimates from individual clusters. The location of clusters is precisely known *a priori*.

# Chapter 5

# Speech Processing Vest

Room reverberations often degrade the speech signal from an active speaker inside a teleconferencing room. Microphone arrays can be of great help in such surroundings in improving the quality of the recorded speech signal. It is very important in such environments to know the actual direction of the speaker. Because of the reverberations, the relative time delays between a pair of microphones may not reveal the correct bearing of the speaker. To combat the effects of the reverberations, distributed pairs of microphones are placed on different locations on the body using a vest. Thus, the goal of the speech-processing vest is to find the direction of arrival of an active speaker in a teleconferencing room, using the weighted DOA's from distributed pairs of microphones. A complete description of the design process of this application is given in this chapter. Issues related to microphone arrays on the human body are discussed in the first section. Maximum likelihood source localization and weighted time-delay estimate based algorithms are described in the following sections. The nature of the simulations performed, the description of the prototype, and the results of the test are discussed in the following sections.

## 5.1 Problem Statement

Speech signals span a range of frequencies from 100 Hz to 5KHz. As explained in Chapter 2, the fundamental formant frequencies range from about 200 Hz to 1000 Hz and vary from person to person. Because of the resonance in the vocal tract, these formant frequencies from voiced speech contain a lot of energy. A vest with a number of microphones distributed on it can be used to estimate the direction of arrival of the speech. The relative time delays between the signals received at different microphones is used to estimate the direction of arrival. Because these formants can cover a decade of frequencies, the underlying direction of arrival estimation algorithm should be able to extract the relative time-delay information from all of these frequencies. Thus, there is a need to use a wideband beamforming algorithm. As explained in Section 2.5, the maximum separation between a pair of microphones should not exceed $\lambda/2$, where $\lambda$ is the wavelength corresponding to the highest frequency of interest. In this case, $\lambda_{min} = \frac{342m/s}{1000Hz} \approx 30cm$ so the maximum separation should be less than $15cm$. The majority of the fundamental formant frequencies for adult human beings lay in the lower range of the spectrum, less than 550Hz. Thus, the maximum separation between the microphone pair can be increased up to $30cm$ in most application.

The primary limitation in using microphone arrays for a wearable vest is that the placements of the microphones have to fit within the dimensions of the human body. The usual small, medium, and large sizes for males have average chest of measurements 91.44, 101.6, and 111.7 cm respectively. The chest size is measured at the fullest part of the body, around the shoulder blades. Thus, the maximum shoulder-to-shoulder separation can vary from about 55.8 cm for the large size to 45.72 cm for a small size shirt. For a large shirt size this allows for the placement of 4 microphones at 13 cm apart and/or 3 microphones at about 25 cm apart. Figure 5.1 shows the possible locations for the placement of the microphones.

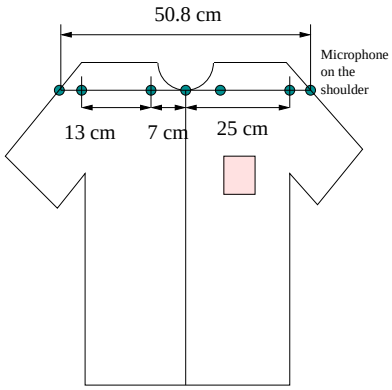


Figure 5.1: Placement of microphones on the shirt

In a small room, the signal emitted from a source is often reflected several times by the walls and the ceiling of the room. The signals seen by the microphones are superimpositions of these reflected, delayed, and attenuated multipaths. Thus, the relative phase delays between all of the microphones may not be the true delays as indicated by Equation 2.5. Depending upon the location of the pair of microphones, the placement of the vest in the room relative to the speaker, and the location of the speaker, different delays are indeed possible. Thus, some of the pairs may see the true delay whereas the others might see many dominant delays. Using the weighted time delay estimates from pairs that have the true time delays, the DOA can be estimated. Hence, a subset of the pairs of microphones that are located on different parts of the body can be used in estimating the true direction of arrival.

Thus, the goal of this application is to look at the advantages of having distributed microphones on the front, the back, and on the shoulders of a wearable vest to estimate the direction of arrival of a distant speaker inside a teleconferencing room. Once the DOA for the active speaker is obtained, the received signals from a subset of microphones can be combined to get a better speech quality.

As explained in Section 3.1, one of the problems with a wearable fabric is that the relative positions of the sensors change with the movement of the body/shoulders or due to the drape

of the fabric. This is especially true for the microphones placed on or near the shoulders. Angular movement of the shoulders and the movement of the torso can change the relative distances between the microphones. Thus it is very important to see the effect of such movement on the performance of the application.

## 5.2 Maximum Likelihood Source Localization (ML-DOA)

In this section, a wideband beamforming algorithm for estimating the DOA of the speech is considered. A maximum likelihood source localization algorithm for wideband signals, presented in [36], is used for this purpose. As explained in Section 2.4, for a single source case, the signals seen by the i$^{th}$ microphone is given by

$$x_i(t) = a_i\, s(t - t_i) + n_i(t) \tag{5.1}$$

$x_i(t)$ : Signal seen at i$^{th}$ the microphone; $i = 0,\ 1,\ 2,\ 3,\ \ldots,\ M-1$

$a_i(t)$ : Attenuation at the i$^{th}$ microphone

$s(t)$ : Signal radiated by the source

$n_i(t)$ : White Gaussian noise

$\tau_i(t)$ :Delay at the i$^{th}$ microphone

$$\text{where} \qquad \tau_i(t) = \frac{||\mathbf{r_s} - \mathbf{r_i}||}{c} \tag{5.2}$$

$\mathbf{r}_s$: Source coordinate

$\mathbf{r}_i$: Coordinate of the i$^{th}$ microphone

A block of L sampled points from the microphone signal can be transformed into the frequency domain by the N-point FFT. A small amount (N-L) of zero padding can be applied to get interpolated points for some of the intermediate frequency bins. Thus, in the frequency domain the array signal can be written in the form

$$X(k) = d(k)S(k) + \eta(k), \tag{5.3}$$

where $k = 0, 1, 2, \ldots, N-1$ and $\eta(k)$ is the zero mean complex white Gaussian noise with variance $\mathrm{L}\sigma^2$ in each element. The spectrum of the individual signals is given by

$$X(k) = [X_0(k), X_2(k), \ldots, X_{M-1}(k)]^T, \tag{5.4}$$

and the scaling matrix is given by

$$d(k) = [a_o e^{\frac{-j2\pi kt_0}{N}}, a_1 e^{\frac{-j2\pi kt_1}{N}}, \ldots, a_{M-1} e^{\frac{-j2\pi kt_{M-1}}{N}}]^T. \tag{5.5}$$

The DOA estimation now involves the search for $\mathbf{r}_s$, which minimises the energy in the difference between $X(k)$ and $d(k)S(k)$ [36]. This optimization criteria for maximum likelihood estimation of the source location and source signal is given by

$$\underset{\tilde{r_s}}{max} J(\tilde{r_s}) = \underset{\tilde{r_s}}{max} \sum_{j=0}^{N/2} ||P(k, \tilde{r_s})X(k)||^2, \tag{5.6}$$

where

$$P(k, \tilde{r_s}) = d(k)d^{\dagger}(k), \qquad \text{and} \tag{5.7}$$

$$d^{\dagger}(k) = (d^H(k)d(k))^{-1}d^H(k). \tag{5.8}$$

For a single source, $P(k, \tilde{r_s})$ can be reduced to

$$P(k, \tilde{r_s}) = \alpha d(k)d^H(k), \tag{5.9}$$

because

$$d^H(k)d(k) = \begin{bmatrix} a_o e^{\frac{j2\pi kt_0}{N}} & a_1 e^{\frac{j2\pi kt_1}{N}} & \dots & a_{M-1} e^{\frac{j2\pi kt_{M-1}}{N}} \end{bmatrix} \begin{bmatrix} a_o e^{\frac{-j2\pi kt_0}{N}} \\ a_1 e^{\frac{-j2\pi kt_1}{N}} \\ \dots \\ a_{M-1} e^{\frac{-j2\pi kt_{M-1}}{N}} \end{bmatrix} \text{ and} \tag{5.10}$$

$$d^H(k)d(k) = a_0^2 + a_1^2 + a_2^2 + \dots a_{M-1}^2. \tag{5.11}$$

For the far-field case, $a_i$ for all the microphones are nearly equal. Hence,

$$d^H(k)d(k) = a_0^2 M = \alpha, \tag{5.12}$$

where $\frac{1}{\alpha}$ is a constant. As long as the source is in the far-field, the search for $\hat{\mathbf{r}}_\mathbf{s}$ can be simplified to a one-dimensional search over the angle of arrival.

## 5.3 Time Delay Estimation Based DOA (TDE-DOA)

This section describes the DOA estimation of the source, based on pairwise weighted DOA estimates. It was shown in Chapter 2 that if the separation between a pair of microphones $\mathrm{d}_{ij}$ is less than $\lambda_{min}/2$, then the relative delay between the signals arriving at the two microphones can be uniquely mapped onto the DOA from either the front or the back,

$$\theta = \arccos \frac{\tau_{ij}\lambda}{d_{ij}}, \qquad \text{where } 0 \leq \theta \leq \pi.$$

Hence, it is necessary to estimate the relative time delays between a pair of microphones. Using N microphones, $\frac{N!}{(N-2)!\,2!}$ pairs can be formed. The time delay estimate from each pair can then be used to compute the same number of DOA estimates. These can then be weighed and a resultant DOA for the source can be calculated. It is well known that the cross correlation $R_{x_1x_2}$, between two signals $\mathrm{x}_1(\mathrm{t})$ and $\mathrm{x}_2(\mathrm{t})$ is related to the cross-power

spectral density function, $G_{x_1x_2}$, by the Fourier transform relationship

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} G_{x_1x_2}(f)e^{j2\pi f\tau}\mathrm{d}f \tag{5.13}$$

where

$$R_{x_1x_2}(\tau) = \frac{1}{T-\tau}\int_{\tau}^{T} x_1(t)x_2(t-\tau)\mathrm{d}t \tag{5.14}$$

$$G_{x_1x_2}(f) = X_1(f)X_2^*(f). \tag{5.15}$$

Here, $X_1(f)$ and $X_2(f)$ are the Fourier transforms of the signals $\mathrm{x_1(t)}$ and $\mathrm{x_2(t)}$. If the signals $X_1(f)$ and $X_2(f)$ are pre-filtered by $H_1(f)$ and $H_2(f)$ to give outputs $Y_1(f)$ and $Y_2(f)$, then the cross-spectrum between the filtered outputs is given by

$$G_{y_1y_2}(f) = H_1(f)H_2^*(f)G_{x_1x_2}(f). \tag{5.16}$$

The generalized cross correlation (GCC) between $\mathrm{x_1(t)}$ and $\mathrm{x_2(t)}$ is thus, given by

$$\hat{R}^{(g)}_{y_1y_2}(\tau) = \int_{-\infty}^{\infty} \psi_g(f)G_{x_1x_2}e^{j2\pi f\tau}\mathrm{d}f, \tag{5.17}$$

where $\psi_g(f) = H_1(f)H_2^*(f)$ [31]. The true time delay estimate is the value of $\tau$ that maximizes $R_{x_1x_2}$. The role of the weighting function $\psi_g(f)$ is to get a sharp peak in $R_{x_1x_2}$ instead of a broad peak so as to get a good time delay resolution. One such weighting method is the phase transform (PHAT) which weights phases irrespective of the SNR's of the frequencies to which it belongs [31]. The GCC-PHAT function gives a sharp peak in the $R_{x_1x_2}$ at the dominant delay of the reverberated signal [3]. Thus, the time delay can be estimated as

$$\hat{\tau} = \underset{\tau}{max}\int_{-\infty}^{\infty} \frac{G_{x_1x_2}(f)}{|G_{x_1x_2}(f)|}e^{j2\pi f\tau}\mathrm{d}f. \tag{5.18}$$

Because PHAT is the GCC method that best suits reverberant environments, this method is chosen for time delay estimation [37]. Once the time delay is estimated, the DOA can be calculated as explained previously.

## 5.4 Weighting of the Time Delay Estimates

Once the time delays for all of the adjacent pairs of microphones are obtained, it is necessary to weight each of these time delays. This is because the estimates themselves may be severely corrupted by the reverberations. Figure 5.2 shows the GCC-PHAT output of signals acquired from two pairs of microphones (ULA) separated by 1 ft in a reverberant room. Figure 5.1 shows the possible locations for the placement of the microphones.

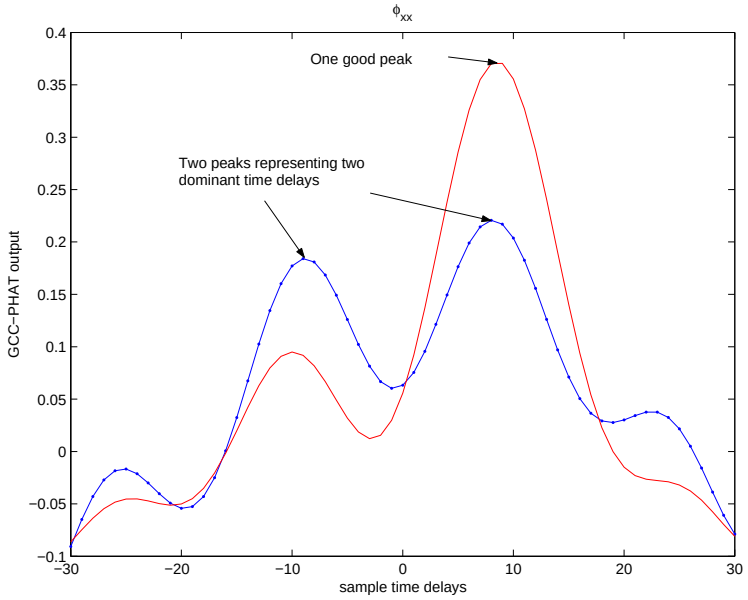


Figure 5.2: Effect of reverberations on GCC-PHAT output

The line corresponding to the first pair (line with markers) in Figure 5.2 exhibits two distinct peaks representing the two separate delays whereas the second pair shows a single distinct peak. It is clear that reverberations degrade the GCC-PHAT output from the first pair. The magnitude of the peaks of the cross correlation output, $R^{(g)}_{x_1x_2}(\tau)$, can be used to get the required confidence in the time delay estimates. Because of the presence of two distinct maximas of comparable magnitude, the confidence level in the time delay given by this pair is low. On the other hand, the delay corresponding to the single largest peak of the second

pair indicates the presence of a single dominant delay. Hence, this time delay ought to be weighted more. The weighting should be based upon the ratio of the difference between the values of the first and the second peak to the value of the first peak of the GCC-PHAT output. Thus, the weighting is given by

$$w_p = 1 - max\left(\frac{c_2}{c_1}, 0\right) \tag{5.19}$$

where $w_p$ is the weighting of the i$^{th}$ pair, $c_1$ and $c_1$ are the magnitudes of the first and the second peaks of the GCC-PHAT output [38]. The maximum value between 0 and $(c_2/c_1)$ is chosen to prevent the assignment of higher weights in the event that value of $c_2$ is negative.

## 5.5 Movement of the Shoulder

The microphones on the shoulder are often displaced from their true locations because of the angular motion of the shoulder. Microphones on the chest are more robust to such displacements. When the new position of the displaced microphones cannot be estimated dynamically, simulations have to be done to characterize the errors resulting from such behavior. In order to see the performance of the vest when a shoulder is displaced or both shoulders are displaced, the following method is used for simulations. When one of the shoulders moves forward, all the microphones on the line joining the end of the shoulder and the center of the chest are assumed to move forward by the same angle. The separation between the microphones is assumed to be unaffected by the rotation. Figure 5.3 shows both of the shoulders moved forward by an angle, $\psi$.

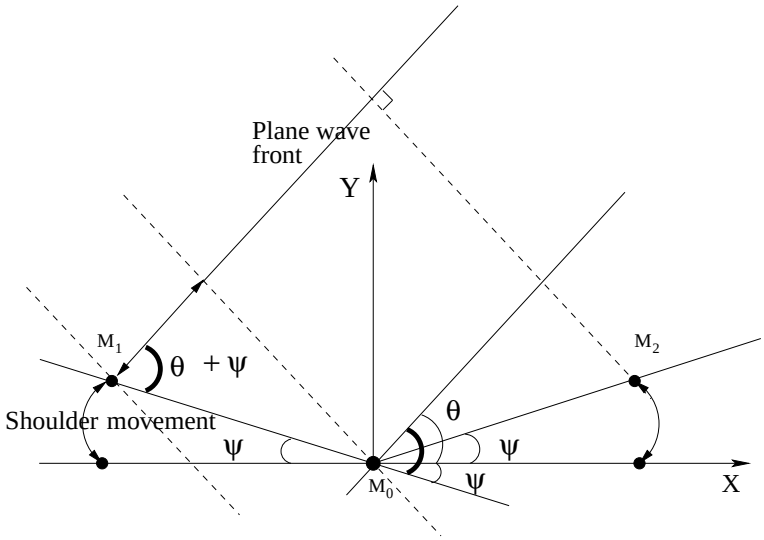


Figure 5.3: Effect of shoulder movement

## 5.6 Simulations in MATLAB

Simulations were done in MATLAB to see the performance of the ML-DOA and TDE-DOA estimation methods. The signal simulated was the superposition of 9 frequencies chosen at random from 200 Hz to 450 Hz. The mean of the difference between these frequencies was 25 Hz. A uniformly distributed random variable between 0 and 1 was scaled by 50 and added to the previous frequency. Phases of these sinusoids at the source were also chosen at random from 0 to $2\pi$. Additive white gaussian noise was added to the signal to have the SNR of the resultant signal as 3 dB. The source was assumed to be in the far field, 30 ft from the center of the microphone array. The angle of incidence was varied from 40 to 140 degrees. The simulations did not incorporate any degradation due to reverberations and attenuation due to the distance travelled. Four different cases were simulated, with each scenario being repeated twice; once with 3 microphones at 1 ft and a second time with 4 microphones with 0.5 ft separation. The first case did not have any shoulder displacement. The second and third scenarios had one of the shoulders displaced by 10 degrees and 20 degrees, respectively.

In the last scenario, both shoulders were displaced by 20 degrees in the forward direction. The results of the simulation can be seen from Table 5.1.

Table 5.1: Simulation results for accuracy of DOA estimation

| **Cases** | **No of mics** | **TDE** | | | **Wideband** | | |
|---|---|---|---|---|---|---|---|
| | | **Abs Avg** | **Std Dev** | **Max Err** | **Abs Ave** | **Std Dev** | **Max Err** |
| **No motion** | 3 | 0.73 | 0.98 | 2.51 | 0.43 | 0.648 | 1 |
| | 4 | 0.53 | 0.62 | 1.53 | 0.3 | 0.56 | 1 |
| **10deg** | 3 | 5.04 | 0.71 | 6.50 | 4.93 | 0.75 | 7 |
| | 4 | 4.93 | 1.03 | 7.3 | 4.93 | 0.6 | 6 |
| **20deg** | 3 | 9.854 | 0.902 | 11.9 | 9.69 | 1.02 | 12 |
| | 4 | 9.893 | 1.069 | 12.52 | 9.83 | 0.735 | 12 |
| **2-20deg** | 3 | 0.68 | 0.88 | 2.17 | 0.77 | 1.30 | 3 |
| | 4 | 1.01 | 1.4 | 3.53 | 1.62 | 2.03 | 4 |

It can be seen that under pure additive white Guassian noise, both ML-DOA and TDE-DOA give almost identical results for all the cases. A decrease in the separation between the microphones increases the mean of the absolute error. It follows from the discussion given in Section 4.3.1 that as the distance between a pair of microphone decreases, the range of values that the sample delays can take also decreases. As the number of available samples delays is finite, the resolution in the delay decreases as the distance decreases. This holds true when the sampling frequency is kept constant for both cases and estimation is done using the shift and add method mentioned in Section 4.2.2. Better resolution can be obtained in the TDE method mentioned in Section 5.3, because the estimation is done in the frequency domain. Intermediate values for the sample delays can be used to find the GCC-PHAT output at

more number of points. In the simulations, intermediate delays were not used, and hence the decrease in the resolution.

The effects of shoulder movement can also be seen from the Table 5.1. When one of the shoulders is moved by 10 degrees, the resultant mean of the absolute error is approximately half the amount of shoulder rotation. This can be explained by a simple 3-microphone array example. Consider that only microphone $M_2$ in Figure 5.3 is displaced by an angle $\psi$. All of the other microphones are assumed to be in their respective positions. For the far-field case, the delay between a pair of microphones $\{M_2, M_0\}$ is given by

$$\tau = \frac{r\ \cos(\theta - \psi)}{c},$$

where $\theta$ is the actual angle of incidence at origin (from Eq( 4.1)) and $\psi$ is the angular displacement of the microphone. Under non-reverberant conditions, the estimated delay will show this value of $\tau$. When the DOA is estimated, the underlying software does not have any knowledge of the angular displacement resulting from the shoulder rotation. Hence, it will calculate the DOA as

$$\theta_{estimated} = \arccos(\frac{\tau\ c}{r}). \tag{5.20}$$

Substituting the value of $\tau$ in the above equation we get

$$\begin{aligned} \theta_{esti} &= \arccos(\frac{\frac{r\cos(\theta-\psi)}{c}\ c}{r}) \\ &= \arccos(\cos(\theta - \psi)) \\ &= \theta - \psi. \end{aligned} \tag{5.21}$$

Hence the error in the estimated DOA is $\theta_{estimated} - \theta = \psi$. The other pair estimates the true delay and hence the true DOA. When both of these DOA's are equally weighted, the average error reduces to $\psi/2$. The analysis can be extended to the four microphones case and the error in estimated DOA will also be $\psi/2$. When both of the shoulders are rotated by the same amount of rotation as shown in Figure 5.3, the pair $\{M_1, M_0\}$ will see $\psi_{1,estimated} = \theta + \psi$

whereas the other pair $\{M_2, M_0\}$ will see $\psi_{2,estimated} = \theta - \psi$. Thus, when both of the DOA's are equally weighed, the resultant DOA will have less error.

## 5.7 Description of the Prototype

The VT e-textiles group, with support from NSF, has built a prototype speech-processing vest. Five microphones are placed on the chest as shown in Figure 5.4. Thin insulated aluminium wires (wrap wires 30 AWG) are sewn into the ready-made large-sized T-shirt. The T-shirt is placed on a dummy made of Styrofoam. Outputs from the omnidirectional microphones are connected to a PCB board shown at the bottom in the Figure 5.4. The board consists of an op-amp amplification circuit with a voltage gain of 80. These amplified outputs are then connected to a data acquisition board, DaqBoard/2000 from IOtech, Inc.. The signals are sampled at 16 KHz sampling rate and quantized with 16-bit precision. Digital signals from the DaqBoard are directly connected to the personal computer using the PCI interface. The personal computer has a Windows 2000 operating system and all the data processing is done using MATLAB 6.1.

## 5.8 Prototype Test Results

The prototype was tested in two different environments. One of the test environments was an auditorium where very few reverberations were present. Another set of readings was taken in a small teleconferencing room.


Figure 5.4: Photograph showing the vest

### 5.8.1 Results from the auditorium

The source was kept at a distance of 10 ft from the vest in the auditorium. A 1 GHz processor with 512 MBytes of RAM processed the data collected by the DaqBoard/2000. The processing was done in MATLAB. A speaker is used to continuously play speech signal, which said, "Primitive tribes have an upbeat attitude" [39]. The fundamental formants of this speech were in the range of 500 Hz. The speaker was moved along a fixed radius to create different angles of incidence. The DOA was estimated using the TDE based approach as well as the ML-DOA approach. Both of these results can be seen from Table 5.2. It can be seen that both of these methods give almost identical performances. Because there were no reverberations in the auditorium, the weights for both of the pairs of the microphones for all the angles of incidence were almost equal. Hence, no performance advantage for the weighting was seen in this case. The mean of the absolute error, the standard deviation, and the maximum error are indicated in the Table 5.6. A maximum of 2 degrees error in the

actual angle could have been recorded because of the error in the physical measurement.

Table 5.2: Readings of DOA estimation using 3 microphones at 1-ft separation in auditorium

| **Actual** | 50 | 56 | 58 | 61 | 65 | 71 | 82 | 90 | 97 | 111 | 115 | 121 | 130 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TDE-DOA** | 48.4 | 52.9 | 54.58 | 57.09 | 63 | 70.68 | 73 | 83 | 89.64 | 102.5 | 109.8 | 111 | 121 |
| **ML-DOA** | 52 | 52 | 55 | 64 | 61 | 70 | 74 | 84 | 89 | 104 | 111 | 112 | 121 |

Table 5.3: Accuracy of DOA estimation in the auditorium

| **Approach** | **No of mics** | **Mean(Abs(Error))** | **Std Dev** | **Max Error** |
|---|---|---|---|---|
| **TDE-DOA** | 3 mics | 4.7 | 3.3 | 10 |
| | 4 mics | 3.01 | 1.55 | 5.41 |
| **ML-DOA** | 3 mics | 5.5 | 2.83 | 9 |
| | 4 mics | 2.22 | 1.645 | 5 |

### 5.8.2 Teleconferencing room results

Figure 5.5 shows the placement of the source and the microphone array in a small teleconferencing room. To evaluate the performance of the application as a function of formant frequencies, DOA estimation was repeated twice, each time with a separate speech source. The room was a moderately reverberant room. The source was kept at a distance of 7 ft from the vest and at the same height as the microphone pair (3.5 ft). The results of the test for the two speech sources are tabulated in Table 5.4 and Table 5.5. As the dimensions of the room are small, it shows moderate reverberations. As discussed in Section 5.4, because of

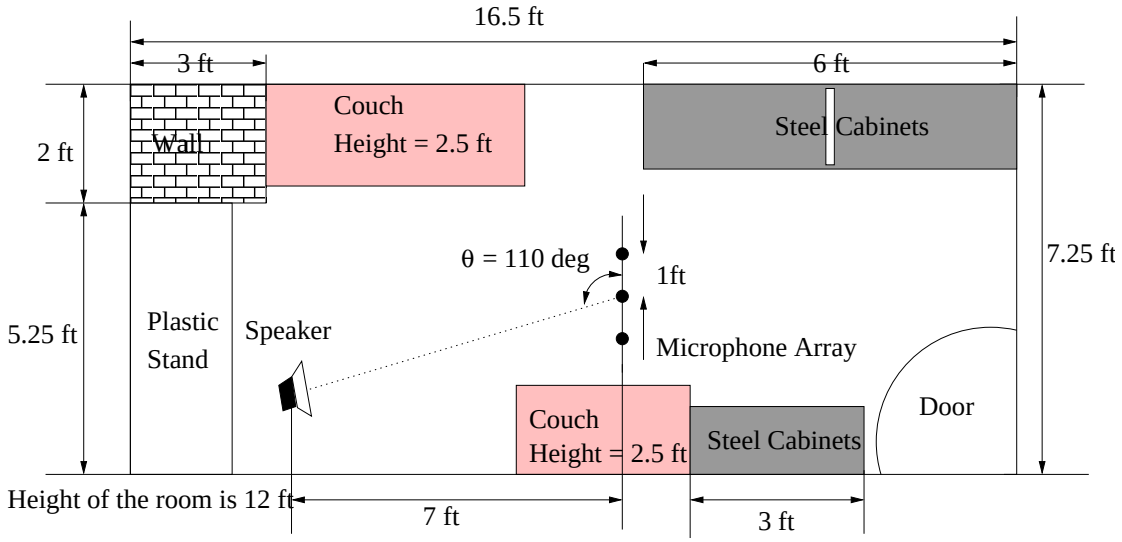


Figure 5.5: Floor plan of the room used for testing

the reverberations, multiple peaks in the GCC-PHAT output can be seen. Figure 5.6 shows a specific extreme case where the GCC-PHAT output from one of the pairs, $\{M_1$ and $M_2\}$, have two dominant delays. In fact, in this case it turns out that the second dominant delay has more power than the power at the true delay. Because the two peaks of this pair have almost the same magnitude, the weight associated with this DOA is very small, 0.0031. On the other hand, the pair $\{M_2, M_3\}$ has one very strong peak corresponding to the true delay estimate. Hence this is weighted heavily and the resultant estimated DOA is 105 degrees. The true direction of arrival is 110 degrees. From these results, it is clear that distributed pairs of microphones on the human body can be used effectively to combat the speech signal degradation due to reverberations.

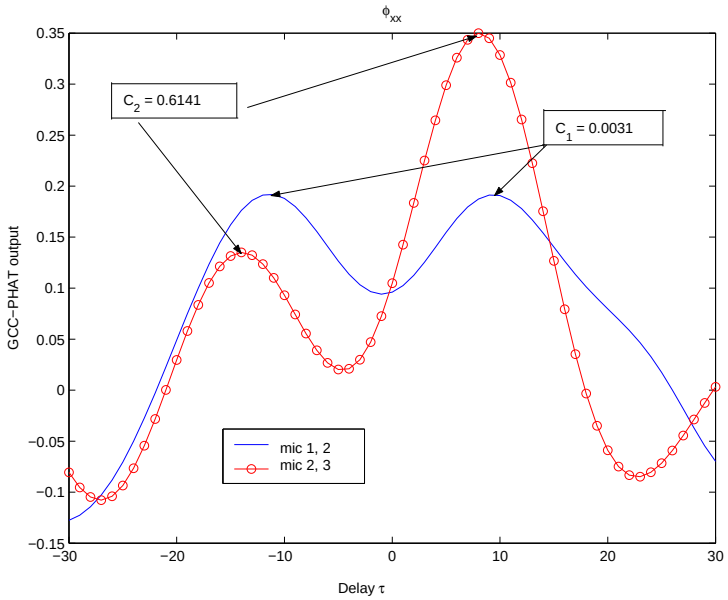


Figure 5.6: Multiple peaks in GCC-PHAT output due to reverberations

Table 5.4: Readings of DOA estimation using 3 microphones in teleconferencing room

| **Actual DOA** | $\tau_{12}$ | **DOA$_{12}$** | $\tau_{23}$ | **DOA$_{23}$** | **w$_{12}$** | **w$_{23}$** | **TDE DOA** | **ML DOA** |
|---|---|---|---|---|---|---|---|---|
| 50 | 7 | 62.24 | 6.5 | 64.37 | 0.256 | 0.182 | 59.12 | 60 |
| 60 | 8 | 57 | 7 | 62.46 | 0.268 | 0.05 | 59.49 | 64 |
| 70 | 3.5 | 76.5 | 3.5 | 76.56 | 0.1 | 0.26 | 76.3 | 72 |
| 80 | 2 | 82.3 | 2.5 | 80.43 | 0.53 | 0.257 | 81.05 | 76 |
| 90 | 0.5 | 88.1 | 0.5 | 88.09 | 0.64 | 0.72 | 88.09 | 82 |
| 100 | -1 | 93.81 | -2 | 97.64 | 0.133 | 0.47 | 96.72 | 89 |
| 110 | -4 | 105.4 | -4.5 | 107.4 | 0.167 | 0.74 | 106.7 | 107 |
| 120 | -6 | 113.5 | -5.5 | 111.45 | 0.256 | 0.096 | 113 | 113 |
| 130 | -8 | 122 | -7 | 117.7 | 0.197 | 0.24 | 119.7 | 120 |

Table 5.5: Readings of DOA estimation using 3 microphones in teleconferencing room

| **Actual DOA** | $\tau_{12}$ | **DOA$_{12}$** | $\tau_{23}$ | **DOA$_{23}$** | **w$_{12}$** | **w$_{23}$** | **TDE DOA** | **ML DOA** |
|---|---|---|---|---|---|---|---|---|
| 50 | 7.5 | 60.06 | 8 | 57.84 | 0.338 | 0.375 | 58.89 | 66 |
| 60 | 5.5 | 68.53 | 6 | 66.46 | 0.512 | 0.619 | 67 | 69 |
| 70 | 3.5 | 76.53 | 4 | 74.56 | 0.302 | 0.537 | 75 | 77 |
| 80 | 1.5 | 84.27 | 2.5 | 80.43 | 0.608 | 0.75 | 82.14 | 80 |
| 90 | 0.5 | 88.09 | -0.5 | 91.90 | 0.696 | 0.578 | 89.82 | 88 |
| 100 | -1 | 93.81 | -1.5 | 95.72 | 0.537 | 0.64 | 94.5 | 94 |
| **110** | +5.5 | 68.52 | -4 | 105.43 | **0.0031** | 0.614 | 105.24 | 114 |
| 120 | -6.5 | 115.6 | -7 | 117.75 | 0.063 | 0.459 | 117.5 | 115 |
| 130 | -8.5 | 124.4 | -9.0 | 125.93 | 0.393 | 0.501 | 125.8 | 120 |

Table 5.6: Accuracy of DOA estimation in the Teleconferencing room

| **Approach)** | **Speech** | **No of mics** | **Mean Abs(Error)** | **Std Dev** | **Max Error** |
|---|---|---|---|---|---|
| **TDE-DOA** | voice | 3 mics | 4.6 | 3.59 | 10.2 |
| | number | 3 mics | 4.23 | 2.47 | 8.9 |
| | voice | 4 mics | 2.944 | 2.55 | 6.54 |
| | number | 4 mics | 3.21 | 2.92 | 9.53 |
| **ML-DOA** | voice | 3 mics | 6.67 | 3.27 | 11 |
| | number | 3 mics | 5.6 | 5.08 | 16 |
| | voice | 4 mics | 4.8 | 2.78 | 9 |
| | number | 4 mics | 3.22 | 2.86 | 9 |

### 5.8.3 Effect of shoulder movement

To see the effects of shoulder movement on the performance of the vest, microphones were displaced as discussed in Section 5.5. The performance of the vest under shoulder movement can be seen Table 5.7. It is evident from the results that the accuracy of the DOA estimate decreases as the displacement of the microphone with respect to its actual position increases. The mean of the average of the absolute error in the DOA estimate is directly proportional to the angular displacement of the microphone. But the standard deviation does not change significantly when only one of the microphones is displaced. This is because of the constant offset in DOA estimate resulting from the angular displacement of the microphone. When both of the microphones are displaced the error in DOA is larger. The standard deviation also increases as the DOA estimate is poorer for incident angles away from broadside.

Table 5.7: Effect of shoulder movement

| **Approach** | **Motion** | **Mean(Abs(Error))** | **Std Dev** | **Max Error** |
|---|---|---|---|---|
| **TDE-DOA** | 1 mic 10 deg | 3.41 | 2.335 | 7.84 |
| | 1 mic 20 deg | 7.18 | 2.91 | 11.02 |
| | 2 mic 20 deg | 7.19 | 4.2 | 13.48 |
| **ML-DOA** | 1 mic 10 deg | 2.55 | 2.4 | 7 |
| | 1 mic 20 deg | 5.055 | 2.871 | 10 |
| | 2 mic 20 deg | 5.55 | 4.719 | 13 |

# Chapter 6

# Conclusion

A framework for designing e-textiles for acoustic applications was presented. A large number of choices for electronic components and textile related options were available. Hence, simulations are indispensable in the design of such applications to explore the complete design space. To verify the function of the application in the real world, constructing a limited number of prototypes is a must. Several iterations of simulation and prototype development have to be done to achieve the desired results. Two applications were designed and the prototypes were tested in physical environments to confirm the functioning of both of the applications.

For the large-scale fabric, the placement of the microphones and the sampling rate are very critical parameters for getting the desired accuracy in the DOA estimation. Higher sampling rates can give significantly better results, but the power consumption increases as the sampling rate is increased. For finding the location of the vehicle using the DOA estimates from multiple clusters, a large separation between the clusters is necessary for getting good results. This is especially true for low sampling rates where the accuracy of the DOA estimation of individual clusters is not very good. Distributing the microphone

arrays on the e-textile can be very useful for real-time speech-processing applications. The spacing between the microphones that are suitable for speech frequencies is readily available in e-textiles. In teleconferencing rooms better DOA estimates can be obtained by using several distributed pairs of microphones on the body. Detailed analysis has to be done to discover the effects of changes in the relative position between the sensors.

## Future research

The field of e-textiles is still in a rudimentary stage. A significant amount of research is needed to address all of the issues mentioned. For the microphone array applications for e-textiles, interconnecting the microphones to the wires in a reliable, cost effective, unobtrusive fashion is of utmost importance. The research can be extended to speech-to-text, speech-recognition, and multi-source separation applications. Independent component analysis and blind source separation algorithms can be used to overcome the issues relating to fixed microphone positions. Directional microphones can also be used for multi-source applications.

# Bibliography


[1] E. Hellweg, "E-Textiles Come into Style: Next season's smart outfits will be wired," in *http://www.technologyreview.com/articles/wo_hellweg080102.as*, $1^{st} of August$ 2002.

[2] J. Edmison, M. Jones, Z. Nakad, and T. Martin, "Using piezoelectric materials for wearable electronic textiles," in *Proc. Sixth International Symposium on Wearable Computers*, pp. 41–48, October 2002.

[3] M. Brandstein, D. Ward, *Microphone Arrays; Signal Processing Techniques and Applications*. Springer-Verlag New York, 2001.

[4] A.P. Pentland, M. Petrazzouli, A. Gerega, T. Starner, "The digital doctor: an experiment in wearable telemedicine," in *Proc. First International Symposium on Wearable Computers*, pp. 173–174, October 1997.

[5] E.J. Lind, R. Eisler, G. Burghart, S. Jayaraman, Sungmee Park, R. Rajamanickam, T. McKee, "A sensate liner for personnel monitoring applications," in *Proc. First International Symposium on Wearable Computers*, pp. 98–105, October 1997.

[6] *Georgia Tech Wearable Motherboard: The Intelligent Garment for the 21st Century*. http://www.gtwm.gatech.edu/.

[7] S. Mann, "Eudaemonic computing ('underwearables')," in *Proc. First International Symposium on Wearable Computers*, pp. 177–178, October 1997.

[8] S. Ram, J. Sharf, "The people sensor: a mobility aid for the visually impaired," in *Proc. Second International Symposium on Wearable Computers*, pp. 166–167, October 1998.

[9] Adrian David Cheok, Fong Siew Wan, Xubo Yang, Wang Weihua, Lee Men Huang, Mark Billinghurst, Hirokazu Kato, "Game-city: A ubiquitous large area multi-interface mixed reality game space for wearable computers," in *Proc. Sixth International Symposium on Wearable Computers*, pp. 156–157, October 2002.

[10] S. Antifakos, B. Schiele, "Bridging the gap between virtual and physical games using wearable sensors," in *Proc. Sixth International Symposium on Wearable Computers*, pp. 139 –140, October 2002.

[11] J. Rantanen, N. Alfthan, J. Impio, T. Karinsalo, M. Malmivaara, R. Matala, M. Makinen, A. Reho, P. Talvenmaa, M. Tasanen, J. Vanhala, "Smart clothing for the arctic environment," in *Proc. Fouth International Symposium on Wearable Computers*, pp. 15–23, October 2000.

[12] K. Kukkonen, T. Vuorela, J. Rantanen, O. Ryyndnen, A. Siffi, J. Vanhala, "The design and implementation of electrically heated clothing," in *Proc. Fifth International Symposium on Wearable Computers*, pp. 180–181, October 2001.

[13] S. Basu, S. Schwartz, A. Pentland , "Wearable phased arrays for sound localization and enhancement," in *Proc. Fourth International Symposium on Wearable Computers*, pp. 103–110, October 2000.

[14] K. Van Laerhoven, A. Schmidt, H.-W. Gellersen, "Multi-sensor context aware clothing," in *Proc. Sixth International Symposium on Wearable Computers*, pp. 49–56, October 2002.

[15] M.M. Gorlick, "Electric suspenders: a fabric power bus and data network for wearable digital devices," in *Proc. Third International Symposium on Wearable Computers*, pp. 114–121, October 1999.

[16] B. Thomas, K. Grimmer, D. Makovec, J. Zucco, B. Gunther, "Determination of placement of a body-attached mouse as a pointing input device for wearable computers," in *Proc. Third International Symposium on Wearable Computers*, pp. 193–194, October 1999.

[17] B. Thomas, S. Tyerman, K. Grimmer , "Evaluation of three input mechanisms for wearable computers," in *Proc. First International Symposium on Wearable Computers*, pp. 2–9, October 1997.

[18] D. Reilly, D. Siewiorek, A. Smailagic, "Power consumption and performance analysis of real-time speech translator smart module," in *Proc. Fourth International Symposium on Wearable Computers*, pp. 25–32, October 2000.

[19] Sung Park, I. Locher, A. Savvides, M.B. Srivastava, A. Chen, R. Muntz, S. Yuen , "Design of a wearable sensor badge for smart kindergarten," in *Proc. Sixth International Symposium on Wearable Computers*, pp. 231–238, October 2002.

[20] J.F. Knight, C. Baber,A. Schwirtz, H.W. Bristow, "The comfort assessment of wearable computers," in *Proc. Sixth International Symposium on Wearable Computers*, pp. 65–72, October 2002.

[21] S. U. Pillai, *Array Signal Processing.* Springer-Verlag, 1989.

[22] Richard. E. Berg, David. G. Stork, *Physics of Sound.* Prentice-Hall, Englewood Cliffs, N.J., 1982.

[23] R.A. Kennedy, T.D. Abhayapala, D.B. Ward , "Broadband nearfield beamforming using a radial beampattern transformation," in *IEEE Transactions on Signal Processing*, vol. 46, pp. 2147–2156, August 1998.

[24] Don. H. Johnson, Dan. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice Hall, Englewood Cliffs, NJ 07632, 1989.

[25] K. Varma, "Time-Delay-Estimate Based Direction-of-Arrival Estimation for Speech in Reverberant Environments," Master's thesis, Virginia Polytechnic Institute and state University, October 2002.

[26] G. E. Pelton, *Voice Processing*. McGraw-Hill series on computer communication, 1993.

[27] H. F. Silverman, S. E. Kirtman, "A two-stage algorithm for determining talker location from a linear microphone-array data," in *Computer, Speech, and Language,*, vol. 6, pp. 129–152, April 1992.

[28] W. Hahn, S. Tretter, "Optimum signal processing for delay-vector estimation in passive signal arrays," in *IEEE Trans. on Information Theory,*, vol. IT-19, pp. 608–614, September 1973.

[29] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, second ed., 1991.

[30] R. O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. PhD thesis, Stanford University Stanford CA, USA, 1981.

[31] C. H. Knapp, G. C. Carter, "The generalized correlation method for estimation of time delay," in *IEEE Trans. on Acoust. Speech Signal Processing*, vol. ASSP-24, pp. 320–327, August 1976.

[32] M. Brandstein, J. Adcock, H. F. Silverman, "A practicle time-delay estimator for localizating speech sources with a microphone array," in *Computer, Speech, and Language,*, vol. 9, pp. 153–169, April 1995.

[33] *Ptolemy II, Heterogeneous concurrent modelling and design in Java.* Department of Electrical Engineering and Computer Sciences, University of California Berckely, August 2002.

[34] R. Riley, S. Thakkar, J. Czarnaski, B. Schott, "Power-aware acoustic beamforming," in *Proc. Fifth International Military Sensing Symposium*, December 2002.

[35] M. T. Jones, T. L. Martin, Z. S. Nakad, R. R. Shenoy, T. Sheikh, D. Lehn, J. Edmison, M. Chandra, "Analyzing the use of e-textiles to improve application performance," 2003. to be published.

[36] J. C. Chen, R. E. Hundon, Kung. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Trans. Signal Processing*, vol. 50, pp. 1843–1854, August 2002.

[37] T. Gustafsson, B.D. Rao, M. Trivedi, "Analysis of time-delay estimation in reverberant environments," in *IEEE Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 2, pp. 2097–2100, 2002.

[38] K. Varma, T. Ikuma, A.A.L. Beex, "Robust TDE-based DOA estimation for compact audio arrays," in *Proc. of Sensor Array and Multichannel Signal Processing Workshop*, pp. 214– 218, August 2002.

[39] "www.cis.ohio-state.edu/ niki/soundemo.html."

# Nomenclature

DOA Direction of arrival

GCC Generalized cross correlation

ML Maximum Likelihood

PHAT Phase transform

SRP Steered Response Power

SRP-PHAT Steered response power-phase transform

TDOA Time difference of arrival

# Vita

Ravi Shenoy was born in Bangalore, India. He did his Bachelors of Engineering [B.E] at Sardar Patel College of Engineering, Mumbai University, with Electronics Engineering as his major in 2001. He joined Virginia Tech in 2001 for his Masters Degree in Electrical Engineering. Since then he has been involved in *e-textiles* research in the Configurable Computing Machines laboratory. Ravi's hobbies include painting, cricket, and tabla (a percussion instrument from India). His technical interests include digital signal processing, digital communications, network performance analysis, and radio frequency integrated circuits.