BINAURAL HEARING PLATFORM FOR FUTURE HEARING AID RESEARCH

Implementation in Matlab Audio Processing framework

ABEL GLADSTONE MANGAM

School of Engineering
Blekinge Institute of Technology
37179 Karlskrona
Sweden
This thesis is submitted to the School of Engineering at Blekinge Institute of Technology in partial fulfilment of the requirements for the degree of Master of Science in Electrical Engineering. The thesis is equivalent to 20 weeks of full time studies.

Contact Information

Author:
Abel Gladstone Mangam
E-mail: abelgladstone@gmail.com
Phone: +46-707 555424

Main Supervisor:
Dr. Nedelko Grbic
School of Engineering (ING)
Blekinge Institute of Technology
Phone: +46 455 385727

Additional Supervisor:
Dr. Bo Schenkman
School of Management (MAM)
Blekinge Institute of Technology
Phone: +46 455 385647

Internet: www.bth.se/ing
Phone: +46 455 385000
SWEDEN
Abstract

Binaural technology is used to recreate realistic sound to the listener. However, many signal processing and noise cancellation techniques do not preserve the binaural characteristics of the binaural sound and thus destroying the realism. Binaural technologies are used in various applications like hearing aids, auditory displays, echo location for the blind, etc. The heart of binaural technology is the Head Related Transfer Functions which provide the transfer functions of the human pinnae, the shoulders etc. These Head Related Transfer Functions provide the Inter aural Level Differences that contributes to the binaural sound. Apart from the Head Related Transfer Functions there are many other cues like the Inter Aural Time differences in binaural hearing that provides information about the nature and location of the source. The Inter Aural Time differences are known to contribute to the localization at low frequencies and the Inter aural Level Differences contribute to localization at high frequencies. This thesis provides a design solution for reproducing the sound field (preserving the binaural cues) with the help of microphone arrays placed on either side of the listener’s ears. The primary work in this thesis is to build a model of a binaural hearing system and implement it in real time using the Matlab Audio Processing framework. This provides a binaural platform for further enhancements like noise reduction and speech enhancement, etc. The system is evaluated mathematically and perceptually. The results show that the platform tracks a single user very accurately and perceptual tests show that the system can be used to localize sounds in the azimuthal plane.

Keywords: Acoustics, Acoustic signal detection Acoustic signal processing, Microphone arrays, Hearing aids, Speaker Localization.
Contents

Abstract i

Contents ii

List of Figures iv

List of Tables vi

Introduction 1

1 Introduction 1
   1.1 The Physiology of Hearing (A Signal Processing Perspective) 2
   1.2 Localization .................................................. 3
   1.3 Interaural Time Differences ................................... 5
   1.4 Interaural Level Differences ................................. 6
   1.5 Duplex Theory of Localization ............................... 6
   1.6 Other cues that help in localization ......................... 7

Design 8

2 Analysis of ITD 8
   2.1 Analysis of the Interaural Time differences ............... 8
   2.2 Modelling ITD in the binaural hearing system .......... 10

3 Analysis of ILD 12
   3.1 Compression of the HRTF ................................. 12
   3.2 Complex Minimax Criterion ............................... 12
Implementation in real time

4 Implementation
4.1 Time Difference of Arrival (TDOA) estimation and far field source location
4.2 Generalized Cross Correlation (GCC)
4.3 Phase Alignment Transform (PHAT)
4.4 Steered Response Power - Phase Alignment Transform (SRP-PHAT)
4.5 Windowed Discrete Fourier Transform (WDFT)
4.6 Results of PHAT in Real Time

Results

5 Methods, Instrumentation and Results
5.1 Construction
5.2 Algorithm
5.3 Testing
5.4 Results
5.5 Analysis of Variance
5.6 Other Conclusions and Observations

6 Conclusion and future work

Appendix

A Matlab Audio Processing (MAP) Framework

Bibliography
List of Figures

1.1 Components of a Binaural Hearing System ................. 2
1.2 The Interaural Coordinate Axis .............................. 4
1.3 Illustration of Interaural Time Differences .................. 5

2.1 Illustration of ITD ........................................... 8
2.2 Plot of ITD for various elevations and the azimuths assuming a head radius $HR = 87.5mm$ and $v = 340.3m/s$ the colorbar indicates the the time difference ............... 10
2.3 Diagram showing the excess path difference because of the introduction of the Ear cups ......................... 11

3.1 Magnitude and phase plots of the Complex Minimax Design with 25% compression .......................... 15
3.2 Magnitude and phase plots of the Complex Minimax Design with 34% compression .......................... 16
3.3 Magnitude and phase plots of the Complex Minimax Design with 54% compression .......................... 17

4.1 Far Field model for estimation of Angle of Arrival ........ 19
4.2 The relationship between the Direction of Arrival vector and the Angle of Arrival .......................... 19
4.3 Model for estimating the TDOA ............................... 20
4.4 The output of the PHAT with different simulated delays with a sampling frequency of 32000 Hz .................... 23
4.5 PHAT in Real time implementation using the MAP framework .......................... 24
4.6 Schematic of the PHAT setup ............................... 24
4.7 Components used for real time testing ........................ 25
4.8 Calculated Angle using the PHAT ............................ 26
4.9 Spline(red) and cubic(blue) interpolation of the angles .... 26

5.1 Left side of Binaural Hearing System ......................... 27
5.2 BHS showing microphones on the Left and the Right Side ... 28
5.3 Microphone bias and the sound card used to interface with the MAP framework ....................... 29
5.4 Block diagram for the implementation of the system using the Matlab Audio Processing framework .................................................. 30
5.5 Frequency response of the BOSE AE2 headphones used. (source: http://www.head-fi.org) ......................................................... 31
5.6 Flow chart of the BHS implementation ........................................ 32
5.7 User Interface for behavioural testing .......................................... 34
5.8 Deviation of location plots for various sources, The radius and the color bar both show the deviation ................................................. 35
5.9 Mean position plot ..................................................................... 36
5.10 The Mean Squared Error of the localization for each user at all the positions ........................................................................... 37
5.11 Bar graph showing the mean opinion score of the ease of locali-
ization ......................................................................................... 38
5.12 Room Impulse response of the test room ..................................... 39
A.1 Schematic showing the MAP framework ..................................... 42
A.2 A sample UI from the MAP framework ...................................... 43
List of Tables

5.1 Analysis of Variance .......................... 37
Chapter 1

Introduction

Binaural signal processing is an important aspect of hearing aids. It refers to the processing of the audio signals without the loss of binaural characteristics in the signal. Intelligent hearing aids further process these signals to reduce noise and provide high quality of speech. However, binaural processing is much more complex when compared to a mono signal, because both the signal (left and right) are correlated with each other and they have more information, like the spatial location, nature of the source, etc.

Most hearing aids process various audio signals at the ear by amplification; pitch shift, etc. so that the persons with impairment will still be able to hear. However, these hearing aids often do not preserve the binaural characteristics. Some hearing aids try to preserve these characteristics by making them as in ear i.e. the hearing aids are inserted into the person’s ear.

However in ear hearing aids are very costly and may not be used by persons other than persons with hearing impairment. Moreover, it is not possible to test new methodologies on such a small device for the cost is very high. The main objective of this thesis is to recreate the effects of spatial hearing in a bigger over the ear type headphone which are inexpensive and to create a system that can be used for testing various hearing aids.

Such a system can also be used for other purposes like noise cancellation devices for people working in industrial areas which have high noise levels. If a speech enhancement system is added it will improve the hearing capabilities of a person in challenging environments with high noise levels like industries.

Figure 1.1 shows the schematic of the components of such system. It must consist of two microphone arrays on either side of the head for estimation of Direction of Arrival (DOA) of the source, a processor for performing various signal processing tasks and a pair of speakers for rendering the sound. The microphone arrays are used for the direction of arrival. Here an assumption is made that the source is in the far field. Only the DOA can
be estimated in the far field. The far field assumption makes the waves arriving at the sensors (microphones) as plane waves thus simplifying the mathematics substantially. The processor is the core of the system design which performs all the signal processing required. For this thesis the processor is the MAP framework that has been developed in [1]. The headphones are used for the reproduction of the sound at the listener's ears. Ideally the headphones must be high fidelity headphones which have an almost linear response throughout the audible frequencies i.e. 20 Hz - 20 kHz. The following chapters discuss in detail about the system and its evaluation.

The problem is to reproduce the sound field at the ears as accurately as possible in real time. The hypothesis in this thesis is that the sound field at the ear can be reproduced fairly accurately by using the Binaural Hearing System. The measure of the accuracy is the ability to localize (to locate) the source when presented to the subjects.

1.1 The Physiology of Hearing (A Signal Processing Perspective)

Human hearing plays a vital role in the localization of sound sources, and we are able to locate very accurately the spatial location just by listening. This information is confirmed by the visual feedback of the eyes. Blind people rely completely on their ears to locate the sources of sound very accurately. There are three main parts of the human ear; the outer ear (auris externa), the middle ear (auris media) and the inner ear (auris interna) [2]. The external ear is a very important part for location determination consists of the pinna or auricular and the external ear canal. The sound that reaches the human ears is transformed by the shape of the pinna and also the human body (nose, shoulders, etc). They act as obstacles to the sound thus causing gain difference (attenuation and resonance) and slow down (time and phase difference). These changes can be described by measuring the spectral changes in the amplitude and the phase of the signal in different spectral bins or components. The spectral changes caused by the head and
torso is found out by first determining the spectrum of the source and then subtracting it from the spectrum of the wave after it enters the ear canal [3]. These spectral differences in the amplitude and the phase describe the sound pressure differences due to the intervening structures (Head, Torso and the Pinnae) can be modelled as a transfer function $H(\cdot)$. Mathematically, using the Fourier transform notation is can be shown as described in Equation 1.1. Where, $y(f, \theta)$ is the spectrum of the wave after it enters the ear canal, $x(f, \theta)$ is the input spectrum both defined as a function of the frequency $f$ and the phase $\theta$. $G(\cdot)$ is the gain.

$$H(f, \theta) = G(f, \theta) \left( \frac{y(f, \theta)}{x(f, \theta)} \right)$$

Here, $H(\cdot)$ is a complex function that is specific for each direction of the wave around each of the ears.

"These transfer functions that describe the changes between the source and the source and the outer ear are called Head Related Transfer Functions (HRTF)" [3].

"HRTFs characterize the transformation of a sound source to the sounds reaching the eardrums, and are central to binaural hearing" [4].

The HRTF is highly specific to a person. It is very difficult to mathematically model a general HRTF that would work with all persons. However, many models are available using wavelets and genetic processing [5][6][7]. One simple model assumes the HRTFs $H(\theta, z)$ to be a function of the angle of the location of the source $\theta$ which can be subdivided into the azimuth and elevation angles. $z$ is the distance of the sound source from the listener. This is called the Common Acoustic Pole Zero Model (CAPZ model). However, this thesis uses a set of standard HRTFs available from the CIPIC (Center for Image Processing and Integrated Computing) database [8]. The reason for using the CIPIC database is that the HRTFs are given in a simple Finite Impulse Response (FIR) filter forms of 200 coefficients for each angle which can be easily searched from the database. However, this thesis proposes a method to compress these HRTFs without any loss of the binaural characteristics.

### 1.2 Localization

Localization refers to the ability of humans to localize a sound source in space. The localization can be done in space assuming the listener's head as the origin. The coordinate system used to denote the angles relative to the listener in space can be called interaural coordinate system, with the line joining the center of the two ears as the interaural axis. It is a spherical coordinate system with the origin at the center of interaural axis. The azimuthal plane is defined as the plane that contains the interaural axis. The
CHAPTER 1. INTRODUCTION

Figure 1.2: The Interaural Coordinate Axis

elevation is defined as the angle made with the azimuthal plane see Figure 1.2. The plane perpendicular to the azimuthal plane is called the frontal plane. The median plane is defined as the plane that divides the head into two symmetrical halves along the nose as shown in Figure 1.2. Localization is performed mainly by the time difference and level difference cues. These cues along with the other cues like the Interaural Phase Differences, etc. are collectively called binaural cues. The salience of the binaural cues depend on many factors such as knowledge of the source characteristics, frequency of the source, the reliability and the plausibility of the cues and consistency of the cues across the spectrum [9].

"It is evident that the perception of locale and attention to a sound source require a transformation relative to the arriving wave front. Since acoustic information arrives at the impinging sound front. Since transformation of the sound is necessary for the sound localization it is reasonable to assume that the external ear served as the acoustical device to perform requisite transformation."[10]
1.3 Interaural Time Differences

Interaural time differences (ITD) are binaural cues that characterize the localization primarily in the azimuthal plane. When the sound reaches the ear from a position in space to the listener, the sound will first reach the ipsilateral ear (the ear that is closest to the source) then after some time delay it will reach the contralateral ear (the ear that is away from the source). This time difference is called the ITD. It is one of the major cues responsible for the localization ability in the azimuthal plane. However, ITD are not unique for all the locations in space. The locus of all the points that have equal ITD forms a cone around each ear, called the cone of confusion. As the ITD arriving from sources that are located in the cone of confusion are the same they cannot be localized uniquely by these cues, they will be differentiated by the use of another set of cues described in the next section. The ITD are dependent on the frequency of sound, as the speed of sound in air is relatively constant. The ITD do not depend on the amplitude. The ITD are used in the low frequencies for localization primarily up to 1500 Hz according to the duplex theory [9]. At higher frequencies the time differences are small and their significance is low. There will be little changes in the ITD of each person due to the anatomical differences. The ITD are illustrated in Figure 1.3

![Figure 1.3: Illustration of Interaural Time Differences](image)
CHAPTER 1. INTRODUCTION

The ITD cues are very dominant below the 1500 Hz frequency, i.e. the low frequencies. [9].

Assuming 90° to be the angle that is completely towards the right of the listener and 0° to be to the right in front of the listener and −90° to be in the left of the listener and 180° to be behind the listener, the ITD increases as the source moves in the azimuthal plane towards the ears and is zero at 90° and 180° respectively. The ITD can be modelled using diffraction theory.

\[
\tau_{lf} = \left(\frac{a}{v}\right) \left(3 \times \sin(\theta)\right) \quad (1.2)
\]

\[
\tau_{lf} = \left(\frac{a}{v}\right) \left(\sin(\theta) + \theta\right) \quad (1.3)
\]

Equation 1.2 is used when the source is located on the ipsilateral side and Equation 1.3 when the source is located on the contralateral side. The equations 1.2 1.3 give the approximate values of ITD (\(\tau_{lf}\)) for a source located at an angle \(\theta\) assuming a spherical head with its center at the origin and the line joining the ears (Interaural Axis) as its axis as shown in figure 1.2. \(a\) is the head radius and \(v\) is the speed of sound.

1.4 Interaural Level Differences

The Interaural Level Differences (ILD) is the other set of binaural cues that are used by people to localize the sound in space. The ILD are more complex and are very hard to model mathematically. The ILD is a cue that will help localize the sources in the median plane. The level differences are primarily due to the presence of pinna and other body parts like the head, neck, torso, etc. The ILD are used for localization at higher frequencies. The ILD are denoted by the HRTF of each ear and are unique for each spatial location.

The ILD cues are frequency dependent and very reliable cues for localization especially at higher frequencies and more importantly in the median plane. If ILD cues are removed from a binaural system it increases the front back confusions. However, there is no definitive evidence that front back confusions are completely eliminated by the ILD cues [9].

1.5 Duplex Theory of Localization

The Duplex theory of localization was proposed by Lord Rayleigh in 1907. According to Feddersen et al. "ITD are used to localise low frequency sounds, in particular, while ILD are used in the localisation of high frequency sound inputs. However, the frequency ranges for which the auditory system can use ITD and ILD significantly overlap, and most natural sounds will have both high and low frequency components, so that the auditory
system will in most cases have to combine information from both ITD and ILD to judge the location of a sound source.”[11]

The duplex theory does not take into account the role of the external ear (pinna) in localization.

1.6 Other cues that help in localization

Apart from the binaural cues there are several other cues that help in localization, these are monaural cues or the Interaural Phase Differences (IPD), the knowledge of the source characteristics, realism and plausibility of the source content, source frequency etc. The next chapter discusses the design and implementation of the two important cues, i.e. ITD and ILD, for designing a binaural hearing system. From here on Binaural Hearing System (BHS) refers to the model of the binaural system being built.
Chapter 2

Analysis of ITD

2.1 Analysis of the Interaural Time differences

The ITD can be designed mathematically with a fair accuracy using simple geometry and construction. Assuming a Spherical head [12] the construction is shown in Figure 2.1.

![Figure 2.1: Illustration of ITD](image)

COH - Centre of the Head in polar coordinates
HR - head radius
L - Distance from the Tangent point
D - Distance from source to COH
DLD - Distance to left ear direct path
CHAPTER 2. ANALYSIS OF ITD

DLA - Distance from the left ear  
DL - total distance to left ear  
DRD - Distance to right ear direct path  
DRA - Distance from the Tangent point to right ear arc length  
DR - Total distance to right ear  
AZ - Azimuth of the source (−180° left to +180° right)  
EL - Elevation of the source (−90° above to +90° below)  
θL - angle from left ear COH to left ear tangent point COH  
θR - angle from right ear COH to right ear tangent point COH  
φ - angle from line D to left and right ears tangent point COH  
AL - angle from the left ear COH to line D  
AR - angle from the right ear COH to line D.  
Xk, Yk, Zk - location of the source in Cartesian coordinates  
v - Speed of sound.

The ITD is defined as in equation 2.1

\[ ITD = \text{Arrival time at the ipsilateral ear} - \text{Arrival time at the contralateral ear} \] (2.1)

Assuming no elevation (EL = 0). From Figure 2.1 L is calculated as

\[ L = \sqrt{D^2 - HR^2} \] (2.2)

\[ \phi = \arccos \frac{HR}{D} \] (2.3)

\[ DLD = \sqrt{D^2 + HR^2 + 2DHR \sin(AZ)} \] (2.4)

\[ DRD = \sqrt{D^2 + HR^2 - 2DHR \sin(AZ)} \] (2.5)

\[ DL = \begin{cases} DLD, & \text{if } DLD < L \\ L + DLA, & \text{otherwise} \end{cases} \] (2.6)

\[ DR = \begin{cases} DRD, & \text{if } DRD < L \\ L + DRA, & \text{otherwise} \end{cases} \] (2.7)

DLA and DRA are given by

\[ DLA = \theta_L HR \] (2.8)

\[ DRA = \theta_R HR \] (2.9)

For the Non Zero elevations the modifications are made to D and AZ only given by Equations 2.9, 2.10
\[ D = \sqrt{X^2 + Y^2 + Z^2} \] (2.10)

\[ AZ = \arcsin (\cos(EL) \cdot \sin(AZ)) \] (2.11)

The ITD are implemented in Matlab for various angles of azimuth and elevation. Figure 2.2 shows the ITD and its dependence on the elevation and the azimuth. It can be observed from Figure 2.2 that the ITD are almost the same for symmetrically placed azimuths and elevations. Moreover it can also be observed that the Elevation angle 90° (median plane) is a place where the ITD are zero for all the azimuths. This model confirms that ITD are not the cues used for localization in the median plane, because the median plane geometrically has zero ITD. However, in reality there are some minor differences in the median plane, and the human ear is very sensitive to very small time differences even in the median plane.

![Figure 2.2: Plot of ITD for various elevations and the azimuths assuming a head radius \(HR = 87.5mm\) and \(v = 340.3m/s\) the colorbar indicates the the time difference](image)

2.2 Modelling ITD in the binaural hearing system

The ITD are not modelled in the binaural hearing system, because the system itself will preserve these binaural cues as it uses two microphones that are approximately placed very close to where the ear canals are located. However, it is very important to understand the models of the ITD, because
the sensors lie outside the ear cups, so there will be a slight change in the geometry of the system as seen in Figure 2.3.

Figure 2.3: Diagram showing the excess path difference because of the introduction of the Ear cups

where \( ab \) are the positions of the ears, \( AB \) are the positions of the microphone arrays. \( OO' \) is the excess path difference that is caused because of the new positions.

Using simple geometry the excess path can be calculated as

\[
OO' = d' \sin(\theta)
\]

In Equation 2.12, \( \theta \) is the angle of the arrival of the source wave. The location of the source is assumed to be in the far field. From the Figure 2.3, \( d' \) is very small compared to the distance \( d \), which is the Inter Aural distance. Hence, it is ignored in the design. One advantage of using the Binaural Hearing System is that the system by its design will have the ITD of the person wearing the system.
Chapter 3

Analysis of ILD

The design of interaural level differences mathematically for the purpose of simulating binaural hearing is a difficult task. The Head Related Transfer functions (HRTF) are the basic functions that are used to transform the electrical signals from the ear phones at the listeners ears. They simply replicate the transformation of the sound as it enters the ear canal through the pinna. These HRTF are highly dependent on the spatial location of the source. The HRTF are complex functions of the azimuth, the elevation, the distance of the source and are dependent on the anthropometry of the persons ears. There has been a lot of research done on interpolating the HRTF. One such method is presented in [13].

These HRTF can be obtained from various databases available on the internet. For this thesis the HRTF has been taken from the CIPIC labs available at [8]. The reason for using this database is that it is well documented and its ease of use with Matlab.

3.1 Compression of the HRTF

The CIPIC HRTF database is a very large database, although this thesis does not try to interpolate the HRTF, it proposes a method for compressing the HRTF using a filter design technique known as Complex Minimax design. This is done to address the limited memory issues when implementing in a Digital Signal Processor (DSP). The Complex Minimax design is used to design a filter which has complex constraints like magnitude, phase and the group delay. The Complex Minimax Design is described in the next section.

3.2 Complex Minimax Criterion

In general the complex minimax criterion can be defined with additional max norm and auxiliary linear constraints as given by Equation 3.1.
\[ \min_{\zeta \in \Omega_1} \max_{\omega \in \Omega_1} v(\omega) |H(\omega) - H_d(\omega)| \]  
subject to, 
\[ \max_{\omega \in \Omega_2} v(\omega) |H(\omega) - H_d(\omega)| \leq \varepsilon_2 \]  
\[ P \xi \leq p \]  

Here, \( H_d(\omega) \) is the desired filter function taken from the HRTF, \( v(\omega) \) is a strictly positive weighting function, \( \varepsilon_2 \) is the max norm bound. \( P \) is \( M \times N \) constraint matrix, and \( p \) is \( M \times 1 \) constraint vector. It is assumed that the domains \( \Omega_1, \Omega_2 \) are closed and all the associated frequency functions are continuous. The above constraints can be rewritten to a continuous semi infinite linear form by using the Real Rotation Theorem.

The Real Rotation Theorem States that:

For a complex number \( Z=x+iy \) the magnitude of the complex number is bounded by
\[ |Z| = \sqrt{x^2 + y^2} = \max_{-0.5 \leq \theta \leq 0.5} \Re\{Ze^{j\theta}\} \]  

The variable \( \theta \) denotes the rotation on the vector, and a set of new constraints are formed for every rotation. E.g. a rectangular approximation have four values of the rotation of the complex vector. Now the complex constraint problem mentioned in Equations 3.1- 3.3. can be written as a semi-infinite linear programming problem.

\[
\begin{align*}
\min \delta \\
v(\omega) &= \Re\{(H_d(\omega) - H(\omega)) \cdot e^{j\theta}\} \leq \delta, \omega \in \Omega_1, \theta \in 2\pi
\\v(\omega) &= \Re\{(H_d(\omega) - H(\omega)) \cdot e^{j\theta}\} \leq \varepsilon_2, \omega \in \Omega_2, \theta \in 2\pi
\\P \xi &\leq p
\end{align*}
\]

Here, \( \theta \) is taken to be an octagonal approximation, i.e. \( \theta \in [0, \frac{\pi}{8}, \frac{3\pi}{8}, ..., \frac{15\pi}{8}, 2\pi] \). And \( V(\omega) = 1 \). The results of the compression are shown in Figures 3.1 - 3.3.

Equation 3.5 can be compactly written in matrix form as
\[ \begin{cases} 
\min \delta \\
A x \leq B 
\end{cases} \]

\( A, x \) and \( B \) can be written as

\[
A = \begin{bmatrix}
\Re\{C\} & -1 \\
\Re\{Ce^{j\frac{\pi}{8}}\} & -1 \\
\Re\{Ce^{j\frac{3\pi}{8}}\} & -1 \\
\vdots & \vdots \\
\Re\{Ce^{j\frac{15\pi}{8}}\} & -1 
\end{bmatrix} \quad B = \begin{bmatrix}
\Re\{De^{\frac{\pi}{8}}\} \\
\Re\{De^{\frac{3\pi}{8}}\} \\
\vdots \\
\Re\{De^{\frac{15\pi}{8}}\} 
\end{bmatrix} \quad x = \begin{bmatrix} h \\ \delta \end{bmatrix}
\]
where,

\[
C = \begin{bmatrix} v_i \phi_1^T \\ \vdots \\ v_i \phi_I^T \end{bmatrix} \quad D = \begin{bmatrix} v_i H_{d1} \\ \vdots \\ v_i H_{dI} \end{bmatrix} \quad \phi = \begin{bmatrix} 1 \\ e^{-j\Omega} \\ e^{-j2\Omega} \\ \vdots \\ e^{-jN-1} \Omega \end{bmatrix} \quad \Omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \vdots \\ \omega_I \end{bmatrix}
\]

Here \( h \) is the desired filter and \( \delta \) is the maximum chebyshev error and \( \{\omega_1, \omega_2, \ldots, \omega_I\} \) are the radial frequencies. Equation 3.6 can be implemented in Matlab using the \texttt{linprog} command, which is used for solving the linear optimization problem.

It can be noticed from Figure 3.2 that the magnitude of the HRTF is matched very well, but there is a certain frequency at which the phase is not matched. The performance of the Complex Minimax method with 25% (Figure 3.3) compression is almost identical to the actual HRTF. The magnitude plots show the magnitude in dB on the y-axis and the the normalized frequency on the x-axis and the phase plots show the phase angle on the y-axis and the normalized frequency on the x-axis.
Figure 3.1: Magnitude and phase plots of the Complex Minimax Design with 25% compression
Figure 3.2: Magnitude and phase plots of the Complex Minimax Design with 34% compression
Figure 3.3: Magnitude and phase plots of the Complex Minimax Design with 54% compression
Chapter 4

Implementation

4.1 Time Difference of Arrival (TDOA) estimation and far field source location

Consider two sensors (microphones) p, q placed at locations $L_p$ and $L_q$, and the line joining these two sensors is called array baseline assuming that the speed of propagation of the sound in the medium is C, and the source is located at a position $S$. Then the TDOA is given by Equation 4.1.

$$T(L_p, L_q, S) = \frac{|L_p - s| - |L_q - s|}{C}$$ (4.1)

The idea of location is to determine $S$ from the Equation 4.1. Given the time delay that is observed between the sensors is $\tau_{(p,q)}$ is known. However, the solution for $S$ from Equation 4.1 is not unique. It is a hyperboloid surface having the sensors at the foci. Therefore the location of $S$ is on a hyperbola for a given TDOA estimate $\tau_{(p,q)}$. However the Angle of Arrival (AOA) can be determined. This is true because of the assumption of far field makes the hyperboloid surface approach the vector that is drawn from the center of the array base line to the source asymptotically. This can be translated in the wave model as: in the far field the wave is assumed to have a planar wave front as compared to a spherical wave front in the near field. By the far field assumption we can rewrite Equation 4.1 as

$$T(L_p, L_q, v) = \frac{L_p^Tv - L_q^Tv}{C}$$ (4.2)

Here $v$ is the unit vector also called the Direction of Arrival (DOA) vector joining the base line and the source. This is related to the AOA as

$$v = [\sin(\alpha) \cos(\alpha)]^T$$ (4.3)

Here $\alpha$ is the AOA. From Equations 4.2 4.3 the TDOA can be written as
\[ T(L_p, L_q, v) = \frac{||L_p - L_q||}{C} \sin(\alpha) \]

Here \( d \sin(\alpha) \) is called the effective sensor spacing, as seen from the direction of the source. Here, it is assumed that there are two sensors and the unit vector \( v \) is two dimensional, hence the source is assumed to be located in a two dimensional plane [1].

Figure 4.1: Far Field model for estimation of Angle of Arrival. [1]

Figure 4.2 shows the relationship between the Direction of Arrival and the Angle of arrival in the far field.

Figure 4.2: The relationship between the Direction of Arrival vector and the Angle of Arrival. [1]

Figure 4.3 shows the modelling of a system for estimating the TDOA between any two sensors. In Figure 4.3 the delays \( \tau_1 \) and \( \tau_2 \) are due to the physical separation of the sensors 1 and 2. It was shown in the equation 4.4. The filter banks are used to convert the signal from time domain to frequency
domain. The correlator is used to correlate the signals from different sensors at various frequencies and then add them. The peak detector detects the maximum of the output of the correlator and estimates the TDOA $\tau_1 \sim \tau_2$ from the estimate of the maximum peak. The correlator used here is the Generalized Cross Correlation (GCC) and the filter banks used is the Windowed Discrete Fourier Transform (WDFT).

![Figure 4.3: Model for estimating the TDOA](image)

4.2 Generalized Cross Correlation (GCC)

The Generalized cross correlation is an extension to the cross correlation, it is achieved by the addition of a specialized gain function, also called, a processor [14]. The cross correlation is given by Equation 4.5

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} G_{xy}(f) e^{2\pi f \tau} \, df$$  \hspace{1cm} (4.5)

Where, $x(t), y(t)$, are the signals from uniform linear microphone array whose positions are adjacent to each other and $G_{xy}(f)$ is the cross power spectrum. The generalized cross correlation between $x(t)$ and $y(t)$ is given by Equation 4.6.

$$R_{xy}^g(\tau) = \int_{-\infty}^{\infty} \psi_g G_{xy}(f) e^{2\pi f \tau} \, df$$  \hspace{1cm} (4.6)

Here $\psi_g$ is given by Equation 4.7

$$\psi_g(f) = H_1(f)H_2^*(f)$$  \hspace{1cm} (4.7)
Here, $\psi$ is called frequency weighting or the processor and $H_1(f)$, $H_2(f)$ are the transfer functions of the sensors and * indicates the complex conjugate. In practice we can calculate only an estimate of the cross power spectrum from finite observations. There are many processors depending upon the application. Here, for localization we use the Phase Alignment Transform (PHAT) processor.

### 4.3 Phase Alignment Transform (PHAT)

The PHAT processor is a processor that is used to whiten the cross power spectrum of two signals. The PHAT processor is given by Equation 4.8 [14].

$$\psi_{PHAT} = \frac{1}{|G_{xy}(f)|}$$  \hspace{1cm} (4.8)

The advantage of the PHAT processor is that it whitens the cross correlation. If the propagation path is not a single path the cross correlation does not have uniform cross power spectrum, instead it will have many dips (poles) which are caused due to reverberation in a real room environment; using a PHAT processor removes the effects of the multi path propagation. It is a very robust processor in reverberant environments.

The GCC-PHAT is used to find the location of the source. The GCC-PHAT is given by Equation 4.9.

$$\hat{\tau}_{xy}(\tau) = \arg \max_{\tau} \int_{-\infty}^{\infty} \psi_{PHAT}(f) e^{2\pi f \tau} df$$  \hspace{1cm} (4.9)

### 4.4 Steered Response Power - Phase Alignment Transform (SRP-PHAT)

The SRP (steered response power) beamformer is given by equation 4.10. The SRP adds the N microphone elements used by aligning their phase, effectively steering the beamformer towards the direction of the maximum power [15]. This can be used along with the PHAT processor to obtain the SRP-PHAT.

$$Y(\omega, q) = \sum_{n=1}^{N} = G_n(\omega)X_n e^{j\omega\Delta_n}$$  \hspace{1cm} (4.10)

$X_n(\omega)$, $G_n(\omega)$ are the Fourier transforms of the $N^{th}$ microphone element and its associated filter response respectively. The equation for SRP-PHAT is given by

$$P(q) = \arg \max_{\Delta_k - \Delta_l} \sum_{l=1}^{N} \sum_{k=1}^{N} \int_{-\infty}^{\infty} \Psi_{lk}(\omega)X_l(\omega)X_k^*(\omega)e^{j\omega(\Delta_k - \Delta_l)} d\omega$$  \hspace{1cm} (4.11)
Here $\Psi_{lk}(\omega) = G_l(\omega)G_k^*(\omega)$ which is analogous to Equation 4.8. Equation 4.11 is an extension of the GCC-PHAT which chooses the GCCs of all the possible microphone pairs and then maximizes the estimate of TDOA using SRP. This improves the resolution of the peak of the true TDOA. The SRP-PHAT is given by Equation 4.11 There are $N$ (when $l = k$) auto correlations in Equation 4.11. These terms contribute to the DC offset of the SRP that is independent of the steering delays.

### 4.5 Windowed Discrete Fourier Transform (WDFT)

The windowed DFT is used to calculate the cross power spectrum in each filter bank in Equation 4.9. The use of WDFT allows us to reduce the effects of spectral leakage. The WDFT is given by Equation 4.12.

$$X \{m, \omega_i\} = \sum_{n=0}^{N-1} W[n] X \{M m + n\} e^{-j \omega_i n} \quad (4.12)$$

Here, $W[n]$ is the windowing function and $\omega_i = 2\pi i/N, i = \{0, 1, 2, \ldots, N - 1\}$, $M$ is the overlap and $m$ is the sub band index. For sake of completeness the IWDFT is also given in Equation 4.13.

$$X \{M m + n\} = \frac{1}{KN} \sum_{i=0}^{N-1} \frac{1}{\pi i} \sum_{k=0}^{k-1} X \{m + k, \omega_i\} e^{j \omega_i n} \quad (4.13)$$

Here, the WDFT is used instead of the usual DTFT to avoid spectral leakage and to have a proper estimate of the Cross Power Spectrum. Further results of the WDFT can be found at [16].

### 4.6 Results of PHAT in Real Time

The performance of the GCC PHAT is simulated and shown below. The delay is simulated by using a delay filter which can be found in [17]. Figure 4.4 shows the simulated results of the PHAT taken from two sensors with varying simulated delays.
Figure 4.4: The output of the PHAT with different simulated delays with a sampling frequency of 32000 Hz

Figure 4.5 shows the implementation of PHAT in the MAP framework which is presented in Appendix A. The speakers are first located at fixed positions they are then switched from one side to the other using an analogue mixer. Both the speakers are identical to each other and white noise is played at each of the speakers. The first speaker is placed at 0° and the other speaker at −30°. The results (calculated angle of arrival) are presented in 4.5.
Figure 4.5: PHAT in Real time implementation using the MAP framework

Figure 4.6 shows the implementation schematic of the above set up. Here 1, 2 are the sensors (microphones), and A, B are the speaker positions. Speaker A is placed at 0° and speaker B is placed at −30°. Figure 4.7 shows the actual set up of the sensor array and the speakers used for the test. The sensors are placed at a distance of 2 cm from each other.

Figure 4.6: Schematic of the PHAT setup

Figure 4.7 shows the microphone array and the Fostex 610B speaker used for the testing of PHAT.
CHAPTER 4. IMPLEMENTATION

(a) Microphone array used for testing the PHAT
(b) Speaker used for testing the PHAT

Figure 4.7: Components used for real time testing

The Figure 4.8 shows the calculated angle using PHAT for various positions of the speaker. It can be noticed that the calculated angle error increases as the source moves towards the end fire regions of the microphone array i.e. towards 90°. This can be corrected using a neural network or hidden markov modelling for limited positions of a source. Figure 4.9 shows the interpolated values for the real data in 4.8 with spline and cubic interpolations using the `interp1` command in Matlab.
Figure 4.8: Calculated Angle using the PHAT

Figure 4.9: Spline(red) and cubic(blue) interpolation of the angles
Chapter 5

Methods, Instrumentation and Results

5.1 Construction

A system was built to preserve the binaural characteristics. The system was then tested both in simulated and real time environments. It consists of a set of hearing protectors, which were modified to fit the microphones that are needed for recording the system. Figure 5.1 shows one side of the BHS.

![Figure 5.1: Left side of Binaural Hearing System.](image)

Each side of the BHS requires an array of at least three microphones for calculating the AOA in space i.e. in three dimensions. The microphones are
AKG C417 PP model with standard XLR connector for phantom powering. These microphones are all taken from a single batch to ensure that the frequency response of the microphones more or less remain the same. The microphones are numbered from 1-6 from left to right as shown in Figure 5.2. Figure 5.3 shows some more pictures of the BHS and its setup.

Figure 5.2: BHS showing microphones on the Left and the Right Side
Figure 5.3: Microphone bias and the sound card used to interface with the MAP framework

The microphones were connected to an external sound card Phase 88 FW Rack which in turn drives the output to the BHS. The sound card supports ASIO (Audio Stream Input/Output) standard. The sound card is controlled via a fire wire cable. The sound card is initiated and controlled by the MAP (Matlab Audio Processing) framework (See Appendix A). The block diagram of the BHS is shown in Figure 5.4.
Figure 5.4: Block diagram for the implementation of the system using the Matlab Audio Processing framework

Each side of the BHS was fitted with an array of three microphones placed in the shape of an L. The horizontal arm of the microphone array is used to find the angle in the azimuthal plane and the vertical arm for finding the angle in the vertical plane. This is the minimal configuration that can be used to find the 3D location of one source in far field. In Figure 5.4 the headphones used are Bose AE2 headphones. These headphones have a very linear frequency response (See Figure 5.5) in the audio frequencies. However, the BHS itself can be fitted with a pair of speakers which can be used for rendering sound in real time. However, this thesis uses the real time recording capability of the MAP framework.
5.2 Algorithm

An algorithm was developed to implement the BHS. A flow chart for the algorithm is shown in the Figure 5.6.
5.3 Testing

Testing the BHS is a challenge, as the system is used to reproduce 3D sound for people. The best way to evaluate and test this system is to have a behavioural hearing test on people. The test was designed in such a way that it can be used for testing the capability of the BHS to preserve the binaural capabilities of the recorded sound, i.e. localization. This test was carried out only in the azimuthal plane. During these tests the record feature of the MAP framework was used. This is done so that the subjects taking the test...
CHAPTER 5. METHODS, INSTRUMENTATION AND RESULTS

will not have any knowledge of the real location of the source as the recording and the testing can be done in different places. It also prevents any leakage that may occur through the headphone caps. This is also done to counter the head movements for which the BHS is not capable to accommodate as it does not use any fixed point of reference. The test can be conducted in real time provided that

• The leakage into the caps of the BHS is very small compared to the signal source.
• The head is very still.

The behavioural test consisted of four different varieties of sound sources; white noise, click train, male speech and female speech. All the sounds are presented to the user at a white noise equivalent of 32 dB SPL A-weighted. All the source sounds were prerecorded on a dummy head with the average head radius of 87.5 mm. The speaker used for recording was a Fostex 6310B (Figure 4.7(b)) is used. The distance from the center of the head to the speaker was 180 cm. The recordings were made in a natural room which was quiet (26 dB(A) SPL). This has been done to allow for room reverberation which makes the sound feel as if it is outside the head [18] [19]. The sound card used with the MAP framework was Terratec Phase 88 Rack FW 24bit/96KHz recording system. The sound card has the capability to record 8 analog microphones and can drive up to 8 sound channels. For the purpose of this experiment 6 input channels (one from each microphone on the BHS) and two output channels (left and right) were used. A set of 12 measurements for each sound source, therefore a total of 48 measurements were made.

There were 19 subjects of which 7 were female subjects. The subjects who took the test were aged between 20 – 39 years with average age being 24 years. None of the subjects had previous experience of taking a hearing test but a few of them reported that they had taken other behavioural tests. All the subjects were given prior instructions about the test. All the subjects were students. None of the subjects reported any hearing impairments, except one (subject 7) who reported a hearing impairment in the high frequencies. The average time each subject took to complete the test was 27 min.

The task of the subject was to find the direction of the sound. However, marking the directions is often a difficult task. To make this easier the recordings were made at 30° intervals and each of the positions are uniquely identified by a corresponding number on a 12 hour clock. This makes it easier for the subject to identify the direction.

A User Interface (UI) was developed to make it easier for the subjects to listen and identify the angles. Figure 5.7 shows the UI for the localization test. The users were given instructions on how to use this UI prior to the
CHAPTER 5. METHODS, INSTRUMENTATION AND RESULTS

34

test. The UI has the capability to provide any of the recorded sources in random order and it also gives the test leader the possibility to store the data provided by each subject in an easy and convenient way for further analysis.

Figure 5.7: User Interface for behavioural testing

5.4 Results

The subjects marked the locations using the UI. However some subjects could not mark the data for a given angle. These skipped directions are omitted from the results.

Figure 5.8 shows the locations plot for all the sound sources. The position plot is a plot of all the locations marked by all the subjects for a given source. This plot shows the deviation for all the positions. The color intensity shows the number of times a position is selected by all the users. The radius of each bubble in Figure 5.8 shows the deviation.
Figure 5.8: Deviation of location plots for various sources. The radius and the color bar both show the deviation.
The mean of the direction as observed by all the users is shown in Figure 5.9.

Figure 5.9: Mean position plot

Figure 5.9 shows that there are visible front-back confusions at 120°, 180°, and 300°. This is probably a result of using a generic set of HRTFs for all the users. 15% of subjects reported that they had no confusions resolving the front and the back. However, one subject quit from the test as he could not feel that he was able to localize any sounds.

Figure 5.10 shows the Mean Squared Error of the positions as found by each subject taking the test, for different types of sources.
Figure 5.10: The Mean Squared Error of the localization for each user at all the positions

One interesting observation can be found from the Figure 5.10. Subject (Subject 7) who reported to have a hearing impairment at higher frequencies was having a significant higher error for the female voice.

5.5 Analysis of Variance

A two way Analysis of Variance was done on the localization data collected from the users to test the significance of the difference in the means of sources, directions and their interaction. Table 5.1 shows the summary of the analysis.

<table>
<thead>
<tr>
<th>sources</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean of Squares</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound Sources</td>
<td>11.46</td>
<td>3</td>
<td>3.819</td>
<td>1.66</td>
<td>0.1746</td>
</tr>
<tr>
<td>Directions</td>
<td>6746.93</td>
<td>11</td>
<td>613.357</td>
<td>266.25</td>
<td>0</td>
</tr>
<tr>
<td>Sources × Directions</td>
<td>81.2</td>
<td>33</td>
<td>2.461</td>
<td>1.07</td>
<td>0.3659</td>
</tr>
<tr>
<td>Error Term</td>
<td>1990.4</td>
<td>864</td>
<td>2.304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8829.99</td>
<td>911</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is evident from the last column in Table 5.1 that for a critical significance value of $p = 0.05$, that there was no significance of the effect of
the sound sources, but the effect of the directions was significant, and the interaction of the sources and the directions was also not significant. This shows that the localization highly depends on the direction of the source rather than the source itself. This also validates that the test designed will only be a test for the localization capability of the subjects.

5.6 Other Conclusions and Observations

Ease of localization as measured by Mean Opinion Score (MOS) was collected from the subjects. The subjects were asked to rate the difficulty of finding the direction for various sources on a scale of 1 - 10 with 10 being the most difficult. The average MOS shows that natural sources are much easier to localize than artificial sources. The MOS scores are presented in Figure 5.11. It can be seen from Figure 5.10 that the error of all the sources is almost the same which is contrary to the MOS scores (See Figure 5.11) where the subjects reported that a natural source was easier to localize. This is probably because of the psychological effects on the localization of a natural source.

Figure 5.11: Bar graph showing the mean opinion score of the ease of localization

Other observations like the precedence effect have also been observed. The precedence effect is defined in [20] as:

"A group of phenomena that are thought to be involved in resolving competition for perception and localization between a direct sound and a reflection."

The precedence of a sound produces a fused image of the direct path and the reflected path of the sound source which will help the listener to localize a source. The BHS has not been psycho-acoustically tested for the precedence
effect. However, the precedence effect can be modelled; a GCC based filter for precedence effect can be found at [19]. Although the recordings were made in a natural room environment, the subjects reported that they could not hear any audible echoes. However, two subjects reported that they were able to hear the echoes when they were presented with the clicks. This shows that the precedence effect is preserved in the BHS. Figure 5.12 shows one of the clicks used on the subjects. It can be seen that the echoes that are recorded are well within the limits of the echo threshold. However the echo is seen to be around 5ms, at this delay the localization dominance of the fused sound image is not effective. The Figure 5.12 is also the room impulse response of the room used for the recording and testing.

Figure 5.12: Room Impulse response of the test room.
Chapter 6

Conclusion and future work

A system was made to build a binaural hearing system using Matlab in real time using the MAP framework. Some basic challenges like addressing the system’s capability to locate a single source, rendering the sources using HRTF have been addressed. However, a lot of work still needs to be done to make the BHS a working and a useful model and more subjective and behavioural tests are required to validate the system. Some of the future works must include the following.

• The building of a speaker on the ear cups. This poses a challenge as the system has to be built in such a way that there are no feedback effects as the microphones are very close to the speaker.

• Addition of more microphones helps to resolve the sources with lesser error. Various constellations other than the linear arrays can also be tested. Calibration of microphones may also be performed.

• Ability to check for the head movements and adjust the HRTF accordingly.

• Ability to remove the room reverberations for improving the speech quality.

• Ability to resolve the front back confusions completely.

• Ability to resolve more than one source and render the sources separately.

• Ability to completely resolve all the possible directions around the head completely and clearly.

• Development of systems like noise cancellation, echo reduction, frequency shifting, etc. to the BHS thus mimicking an intelligent hearing aid.
CHAPTER 6. CONCLUSION AND FUTURE WORK

- Implementation of the BHS in a DSP processor to make a standalone system without the use of the computer.

- To study the perceptual effects on localization in a quantitative manner.
Appendix A

Matlab Audio Processing (MAP) Framework

The MAP (MATLAB audio Processing tool) framework is a tool that is developed by the acoustic research group at Blekinge Institute of Technology [1]. It is a data acquisition tool that is used to get a data from a sound card that supports the ASIO system in real time. The framework provides an interface between the Matlab environment, the user and the real time environment.

![Schematic showing the MAP framework.](image)

Figure A.1: Schematic showing the MAP framework. [1]
APPENDIX A. MATLAB AUDIO PROCESSING (MAP) FRAMEWORK

The Figure A.1 shows the MAP framework. The framework takes the input from the Audio driver. (Note: The audio driver must support the ASIO standard to work with the framework). Matlab environment is used to write a user defined script that changes the signal from the input and then this signal is fed to the output as shown in Equation A.1. All the processing is done in blocks.

\[ X_n = f(y_n[k]) \]  \hspace{1cm} (A.1)

In Equation A.1 \( k \) represents the length of the block or the buffer length, \( n \) is the frame number. The function \( f(.) \) is the user defined script that is used to process the input signal.

The major advantage of using the Matlab framework is that it provides the user with a very clean interface that is free from setting up or installing all the user would need is a sound card that supports the ASIO standard. The ASIO standard is a standard that is developed to bypass the operating system kernel mixer so that it provides a low latency interface. The MAP framework has a UI that is used to give the information to the user about the dropped blocks or the usage of the processor. Figure A.2 shows the UI example of the MAP framework.

Figure A.2: A sample UI from the MAP framework

More information about the MAP framework and its evaluation can be found in [1]
Bibliography


