A DEVICE TO LOCALIZE SOUND SOURCES

by

CAROL PATRICIA HARRIS

B.A.Sc., Queen's University, 1992

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Electrical Engineering

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July 1995

© Carol Patricia Harris, 1995.

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of ELECTRICAL ENGINEERING

The University of British Columbia
Vancouver, Canada

Date JULY 28 / 1995

DE-6 (2/88)

ABSTRACT

This thesis contains a detailed report of the design, construction and testing of the prototype of a new assistive listening device. The project objective was to design a device that would use a sound localizing algorithm based on human auditory function to determine the position of a target sound source, and would then automatically steer a highly directional microphone to that position. Preferentially amplifying only the intended sound source provides a much cleaner signal for the user, and improves speech intelligibility. Such a device would be useful in controlled environments such as in the classroom, at a business meeting, or situation where there was orderly conversation with only one speaker at a time.

The completed device consists of an array of three omnidirectional microphones, a small directional microphone mounted on a motorized rotating platform, and a digital signal processing (DSP) board. Sound samples collected from the omnidirectional microphones are stored on the DSP board, and are analyzed to determine the position of any sound sources present. If a source is identified, the directional microphone is steered to point in the direction of the source. The directional microphone has a range of travel of 240 degrees, through which it can rotate in under one second, and settle on the chosen location with an error of less than five degrees. The performance of the device as a whole was tested in an anechoic environment. Results showed that the device could localize voice, sinusoidal, and pulsed signals with average errors over a variety of test configurations ranging from approximately three to six degrees.

TABLE OF CONTENTS

Abstract ii

Table of Contents iii

List of Tables vi

List of Figures vii

List of Abbreviations ix

Acknowledgements x

Chapter 1: Introduction 1

1.1 Background 1

1.2 Review of Hearing Aids and Assistive Listening Devices 3

1.2.1 Hearing Aids 3
1.2.2 Assistive Listening Devices 5

1.3 A New Approach 8

1.4 Motivation, Project Objectives, and Scope 11

1.5 Thesis Outline 12

Chapter 2: Theory 14

2.1 Introduction 14

2.2 Binaural Hearing 14

2.3 Properties of Speech 18

2.4 Sound Localizing Algorithm 21

2.5 Signal Processing 28

2.5.1 Fourier Transforms 28
2.5.2 Convolution and Digital Filtering 30
2.5.3 Cross-Correlation 32

2.6 Control Theory 33

2.6.1 The Mechanical System 33
2.6.2 Feedback 36
2.6.3 Parameter Definitions 37
2.6.4 Starting and Stopping 40

Chapter 3: Implementation 41

3.1 Introduction 41

3.2 Design Specifications 41

3.2.1 Stage 1 -- Data Input Design Specifications 41
3.2.2 Stage 2 -- Sound Localizing Algorithm Design Specifications 43
3.2.3 Stage 3 -- Data Transfer Design Specifications 44
3.2.4 Stage 4 -- Directional Microphone Positioning System Design Specifications 45

3.3 Mechanical Considerations 47

3.3.1 Sound Localizing Algorithm Resolution and Mechanical Design 48
3.3.2 Silent Operation and Mechanical Design 55
3.3.3 Construction Details 58

3.4 Electronics 66

3.5 Signal Processing 70

3.5.1 Data Input 71
3.5.2 Sound Localizing Algorithm 73
3.5.3 Control Algorithm 86

Chapter 4: Experiments and Results 93

4.1 ASLAD Specifications 93

4.1.1 Stage 1 -- Input Stage 93
4.1.2 Stage 2 -- Sound Localizing Algorithm 95
4.1.3 Stage 3 -- Data Transfer - Sound Localizing Algorithm to Directional Microphone Positioning System 96
4.1.4 Stage 4 -- Directional Microphone Positioning System 97

4.2 Performance Evaluation in an Anechoic Environment 101

4.2.1 General Observations 105
4.2.2 Errors in Sound Localization 108
4.2.3 Performance in Noise 112
4.2.4 Time Required to Locate and Move to a Target 116
4.2.5 Behaviour when Presented with Two Targets Simultaneously 118

4.3 Qualitative Evaluation of the Sound Localizing Device in an Uncontrolled Acoustic Environment 119

Chapter 5: Discussion and Conclusions 121

5.1 Discussion and Recommendations for Future Work 121
5.2 Summary and Conclusions 127

References 131

LIST OF TABLES

3.1	Design specifications sheet for ASLAD.	42
3.2	Look-up table for determination of target range.	83
4.1	Specifications sheet for ASLAD.	94
4.2	Mean error and standard deviation for sound localization experiments with AM sine signal.	109
4.3	Mean error and standard deviation for sound localization experiments with live voice.	109
4.4	Time required to locate and move to a new target.	109

LIST OF FIGURES

1.	Flowchart of conceptual design -- the necessary components of the assistive listening device.	10
2.	Origin of the differences in sound at each ear from a single target sound.	16
3.	Spectral and temporal characteristics of speech and noise.	19
4.	Definition of variables for derivation of target angle and localization resolution for two omnidirectional microphones.	24
5.	Maximum localization resolution as a function of target position and sampling rate.	27
6.	Basic circuit diagram for a motor and load.	34
7.	Definition of parameters used to characterize a typical second order system.	38
8.	Three omnidirectional microphone design.	52
9.	Four omnidirectional microphone design.	54
10.	Maximum localization resolution as a function of position for the three microphone design.	56
11.	Dimensioned drawings of ASLAD.	59
12.	Side view of the internal framework of the directional microphone positioning system.	63
13.	Gain characteristics of the Phonak SDM directional microphone.	65
14.	Circuits used in ASLAD.	67
15.	Flowchart of the sound localizing algorithm.	74
16.	Digital low-pass filter used to obtain the envelope of the recorded sound samples.	75
17.	Flowchart of the desired position algorithm.	80

18. Flowchart of the control algorithm for the directional microphone positioning system. 90

19. Position profile for rotation of directional microphone through different angles. 99

20. Examples of control and position data. 102

21. Experimental setup in the anechoic chamber. 104

22. Noise signals used in testing localization ability of ASLAD in the presence of noise. 114

LIST OF ABBREVIATIONS

A/D	--	Analog-to-Digital
AM	--	Amplitude Modulated (or Amplitude Modulation)
ASLAD	--	Automatic Sound Localizing and Amplification Device
D/A	--	Digital-to-Analog
DMPS	--	Directional Microphone Positioning System
DSP	--	Digital Signal Processing (or Processor)
FFT	--	Fast Fourier Transform
IFFT	--	Inverse Fast Fourier Transform
SPL	--	Sound Pressure Level

ACKNOWLEDGEMENTS

Thank you to my family and my husband for their support and encouragement.

I would like to thank Drs. C.A. Laszlo, M.S. Cynader, and P. Zakarauskas for the opportunity to work on this project, and for their help and advice throughout.

I would like to thank Dr. W.G. Dunford for taking the time to discuss various aspects of the project with me. In particular, discussions about driving systems for the directional microphone and discussions on future work that could be done to improve the overall performance of the system were much appreciated. Thanks also to Dr. M. Yedlin for many discussions on digital signal processing and for his general encouragement. I would like to thank Dr. M. Hodgson of the Mechanical Engineering department for the use of the anechoic chamber. Finally, thank you to K. Dotto for helping me debug my program.

This research was supported by the IRIS network of Centres of Excellence.

CHAPTER 1: Introduction

1.1 Background

It is estimated that seven percent of Canadians suffer from some degree of hearing impairment (Shein, 1992). This number is probably similar for most developed countries worldwide (Davis, 1989). Hearing ability often degrades with time (Van De Graff, 1988), so as the average lifespan increases, this number is likely to rise. A hard of hearing individual is defined as someone with any degree of hearing impairment who primarily communicates through hearing and speech. Most commonly, but not exclusively, this is a person who has had normal hearing, and has suffered loss as a result of age, illness, injury, or damage due to prolonged exposure to noise.

There are many types of devices designed to help the hard of hearing individual interact with others. They can be divided into five basic categories: hearing aids, assistive listening devices, visual aids, alarm systems, and telecommunication systems. The traditional hearing aid is designed to fit in or behind the ear, or somewhere on the body. Assistive listening devices may be hand held, worn on the body, or placed at any suitable spot close to the speaker(s). This type of device generally incorporates a microphone and a processing system that performs the functions of amplification and noise reduction. The output of such a device is transmitted through a suitable receiver to either an existing hearing aid or to some other type of sound transducer. A new portable assistive listening device, designed to sit on a table in front of the user is the subject of this thesis.

Difficulty in understanding speech in the presence of other sounds is the most common complaint of hard of hearing people (Sammeth and Ochs, 1991). These other sounds may be interfering noises or competing speakers. One of the most common problems with hearing aids

is that while sound can easily be amplified so that it can be heard by the user, very often the desired sounds and unwanted noise are equally amplified. The result is that, for the hearing aid user, speech tends to become blurred and difficult to interpret. The unique acoustical cues that a person with normal hearing ability uses to separate and interpret competing sound sources are not available to the hard of hearing person. If there is additional amplified noise mixed with speech it becomes very difficult, and often impossible, for the hard of hearing person to understand what is being said. If the noise source is predominant, it can mask any desired signal to the point where no useful information is obtained by the hearing aid user. For the hard of hearing person, the effect of noise mixed in with desired speech is that it takes longer to process the sounds being heard and figure out what was actually said. The delay in the processing of speech often makes it difficult to take part in group conversations, as opportunities for participation are lost due to slow response time.

Test results reported by (Plomp, 1994) show that in a situation where a listener was able to understand 90% of a set of test sentences at a given speech to noise ratio, an increase of 1 dB in noise level resulted in a 15% decrease in intelligibility scores. Similarly, a 3 dB increase in the noise level resulted in a 50% decrease in intelligibility scores, and a 6 dB increase resulted in an 80% decrease in intelligibility scores. Therefore, it is important to make the signal received by the hard of hearing person as clear of interference and distortion as possible. Thus, the rationale for using assistive listening devices is to obtain the "cleanest" signal, with the highest signal to noise ratio, and transmit that signal to the hard of hearing person for listening with or without a hearing aid.

1.2 Review of Hearing Aids and Assistive Listening Devices

1.2.1 Hearing Aids

The basic function of a hearing aid is the amplification of sound to a level that can be heard by the user. There are many different types of hearing aids on the market, aiming to satisfy the varying needs of hard of hearing people. One of the common problems among them is that unwanted noise and desired sounds are equally amplified. To this end, most hearing aids employ some type of signal processing scheme for the purposes of noise reduction. The most commonly used method of noise reduction is filtering of the incoming signal in an attempt to reduce noise in frequency ranges not likely to contain speech information. For example, many common noises, such as those created by the ventilation system of a building, have substantial low frequency power. This tends to mask the higher frequency components of speech. High pass filtering will remove much of the low frequency noise, while leaving the majority of the speech signal intact (Levitt *et al.*, 1989). This type of filtering has little effect in the case of competing speech sources. Another method that is sometimes used is noise cancellation. This can be done using either one or two microphones. Noise cancellation using only one microphone requires a constant sampling of the incoming signal to determine if it contains speech and noise, or noise only. From the sampled intervals that are deemed not to contain speech, an estimate of the spectral characteristics of the noise is obtained. This estimate is constantly updated to allow for changes in the acoustic environment, and is subtracted from the spectrum of the speech plus noise signal (Levitt *et al.*, 1989). Again, in the event that a good estimate of the noise cannot be obtained, or if the interfering sound is a competing speaker, this method may not provide the desired result. Another noise reduction method that has been tested for use in hearing aids is to perform noise cancellation by using an omnidirectional microphone to collect the primary signal, and a directional microphone, mounted pointing backwards, to collect a secondary signal. The

directional microphone is assumed to be picking up unwanted noise, and thus its signal is subtracted from the forward facing omnidirectional microphone (Schwander and Levitt, 1987; and Weiss, 1987). This method can make improvements to the sound quality, but it requires the user to constantly change position to obtain maximum benefit.

Many hearing aids amplify sound incident from all directions. The amplification of all sounds means that in a noisy environment, where several sound sources are present, the hearing aid user is unable to make sense of any of them, as they are all equally amplified, summed and presented to the eardrum. As described above, the ability of a hard of hearing individual to separate different sounds is compromised. At present, there is no suitable method for selectively choosing and amplifying only one source that could be implemented in a hearing aid (Engebretson, 1994).

In addition to noise reduction, some hearing aids also incorporate further processing that is intended to shape the incoming signal to match the hearing ability of the user over most frequency ranges. An example of such a process is the use of compression techniques, which limit the dynamic range in specific frequency ranges to fall above the threshold of audibility but below the threshold of pain (Sammeth and Ochs, 1991). Unfortunately, compression techniques can also distort speech in ways that may negatively affect speech recognition abilities (Engebretson, 1994).

In most cases, the nature of the hearing loss and personal preferences dictate what sort of filtering or other processing is the most effective. Almost all hearing aids are custom tuned to maximize the understanding of speech by the individual. Many have multiple filter settings, any one of which can be selected by the user to give the best results for the situation at hand. To date, there has been little success in producing a hearing aid that uses true digital signal processing. Since there are many gaps in our knowledge about the physiological processes of

the auditory system, it is unknown what digital signal processing can successfully be applied to improve the speech discrimination of hard of hearing people (Engebretson, 1994).

1.2.2 Assistive Listening Devices

In order to reduce the effect of acoustical distortion and noise, assistive listening devices are used with or without hearing aids. Such devices generally shorten the acoustical pathway between speaker and listener by placing a microphone close to the speaker's mouth. Transmission to the listener is accomplished using modulated infrared, FM or magnetic coupling.

An alternative approach is to use some type of multi-microphone system. Some systems use a directional microphone to acquire the desired speech signal, and an omnidirectional microphone, mounted at an angle of 180 degrees relative to the directional microphone, to record background noise. The signals from both microphones are then applied to an adaptive noise canceller in an attempt to subtract the background noise from the primary signal (Weiss, 1987). For this approach to be effective, the user must continually aim the primary microphone in the desired direction. Other systems use a second microphone placed at some distance from the primary microphone to record background noise. The output of the secondary microphone is then again subtracted from the primary microphone (Feder *et al.*, 1989). Many different attempts have been made to apply adaptive beamforming and adaptive filtering algorithms to a pair of microphones placed either side by side or one behind the other (e.g. Greenberg and Zurek, 1992; Schwander and Levitt, 1987). With these two-microphone systems, tests have shown that an improvement in speech intelligibility can be obtained in controlled environments where there is a single noise source with significant spatial separation from the speech source. It is thought that the performance of two microphone systems will decrease with the addition of multiple noise sources and the allowance of head movement by the user (Soede *et al.*, 1993a).

In recent years, there have been a number of publications on the development of portable microphone arrays of anywhere from three to seven sensors, suitable for assistive listening devices (eg. Soede *et al.*, 1993a; Kates, 1993; Hoffman *et al.*, 1994). It is thought that a small array of microphones could be incorporated into the arm of, or across the top of the frame of, a pair of eyeglasses. These small arrays are used to provide directional selectivity by weighing the inputs from each microphone in the array and summing the weighted inputs to preferentially amplify sounds incident from a particular direction. This is known as beamforming, and it is essentially a spatial filter. There are two types of signal processing schemes for spatial filters: fixed and adaptive.

In fixed signal processing, the weighting functions for the signals from each sensor in the microphone array are kept constant, producing a fixed directional pattern. These are generally designed such that sources directly in front of the listener will be optimally amplified. The results of these small arrays have been good. For example, Soede *et al.* (1993b) report a mean improvement of the speech reception threshold in noise (level at which 50% of speech presented is correctly identified) of 7.0 dB for a five microphone array mounted across the front of a pair of eyeglasses, and a 6.8 dB improvement for a five microphone array mounted along the arm of a pair of eyeglasses. This improvement was measured as compared to speech reception thresholds in noise with a single omnidirectional hearing aid microphone. With a fixed array, the user has to be looking directly at the source that s/he wishes to listen to. This can limit the user's freedom to move about. For instance, in a business or classroom environment, the user would not be able to look down at a note book or take notes without losing the signal from the desired speaker. In addition, the user must first know where the noise source is coming from in order to be able to look at it. For the hard of hearing individual, this is not always trivial. Freedom of movement can be restored to the user by using the directional array as a hand held

or table top device, but again, the user must then be able to identify the position of each new desired sound source, and constantly reposition the array for optimal gain of that desired source.

In adaptive signal processing for microphone arrays, the contribution to the final signal from each microphone is continuously updated based on information obtained from the incoming sound signals. In theory, the shape of the spatial filter can continuously be adjusted to accommodate changes in the sound environment. The filter can be reshaped for optimal gain in the direction of a desired source and low sensitivity in the direction of unwanted noise sources. The processing required for adaptive arrays is generally more intensive. This is because the characteristics of the incoming sound sources must constantly be analyzed to determine the location of speech and noise sources, and new weighting factors for each microphone in the array must be calculated. If worn on the body, the microphone array would likely be in constant motion, complicating the process of finding desired sources and undesired noise locations. For this reason, an adaptive spatial filter is more suitable for a stationary device. For adaptive spatial filtering, precise information about the microphone array geometry and other array parameters is necessary to achieve good results. Slight misalignment of the array elements and missteering of the array due to inaccurate calculation of the desired source position can cause cancellation of the desired signal in excess of 30 dB for some frequencies (Hoffman *et al.*, 1994). At present, the results for adaptive multi-microphone arrays are not as good as the result achieved by manual positioning of a fixed array (Link *et al.*, 1992).

Systems using large two dimensional arrays of microphones have been reported in the literature (Flanagan *et al.*, 1985). Beamforming algorithms are used to focus the array on a desired source, and preferentially amplify it over other sounds. These systems work well, but are bulky because of the dimensions of the microphone array and hence not appropriate for hearing aid applications or personal assistive listening devices. They are intended for use in

conference facilities or lecture halls. In addition to size, the focussing on an identified sound source while simultaneously scanning for new sources is computationally very intensive for large arrays. The system developed by Flanagan *et al.* (1985), consists of a 9 by 7 array of microphones. This system can provide spatial discrimination over the range of frequencies from 200 to 3400 Hz, and is capable of steering a beam over a 60° range.

1.3 A New Approach

The use of spatial filtering has been shown to provide improvement in the speech reception threshold in noise for hard of hearing individuals. However, adaptive array techniques have not yet proven to be reliable for assistive listening device applications. The fixed array requires constant positioning and, in the case of eyeglass mounted devices, may restrict the movement of the user. The user must also be able to locate sound sources independently. In order to have a personal assistive listening device that successfully makes use of spatial filtering to improve the speech reception threshold in noise, the following is necessary: a way to find the location of a sound source, and a way to preferentially amplify the sounds from the identified location. Ideally, the device should also be small enough to be portable, it must be quiet, it must operate in real time, and it should be automatic to allow hands-free operation. Most importantly, it must make an improvement in the user's ability to understand speech in noise. By considering separately the functions of localization and amplification, instead of trying to perform both with the same instrumentation, we have developed a new method for automated sound localization and amplification.

For a method of sound localization we turn to the human auditory system. In recent years, there has been a growing interest in the development of sound localization methods based on human (or animal) auditory function. Sound localizing algorithms have been developed using

both traditional signal processing techniques (eg. Bhadkamkar and Fowler, 1993), as well as the relatively new neural network (eg. Yuhas, 1992; Palmieri *et al.*, 1992). Humans possessing normal hearing function exhibit excellent sound localizing ability with only two sensors (ears). In an anechoic chamber, humans can identify the location of single sound sources to within one degree horizontally and five degrees vertically if the source is in front of the subject (Mills, 1958), and to within five degrees horizontally if the source is to one side of the subject (Colburn, in press). Research in the field of auditory localization has revealed basic cues that are believed to be used by humans to localize sound sources. Studies have also shown that the speech reception threshold for humans is as much as 10 dB lower (better) for binaural hearing over monaural hearing in cases where the primary and interfering sources are at different locations in space (Duquesnoy, 1983). However, it is not known precisely how and in what combinations these cues are used to identify and separate competing sound sources. So while it may not yet be possible to use a model of the human auditory system to amplify one identified source over another, we may be able to use it to give localization information. Our ears are separated by a distance of approximately 20 cm. A system that can provide position information with only two sensors at a separation of 20 cm is a good candidate for a personal assistive listening device. This would help to keep both the size and the processing requirements to a minimum.

A simple way to aim a directional microphone or fixed directional microphone array would be to move it mechanically. Figure 1 is an information flow diagram showing the breakdown of major components required in a device that localizes sound sources and steers a directional microphone towards the identified source. The process is divided into four stages. Stages two and four are associated with the principal components of sound localization and directional microphone positioning respectively. The stages of data input and data transfer are shown explicitly because of the importance of the performance of these processes. Data transfer

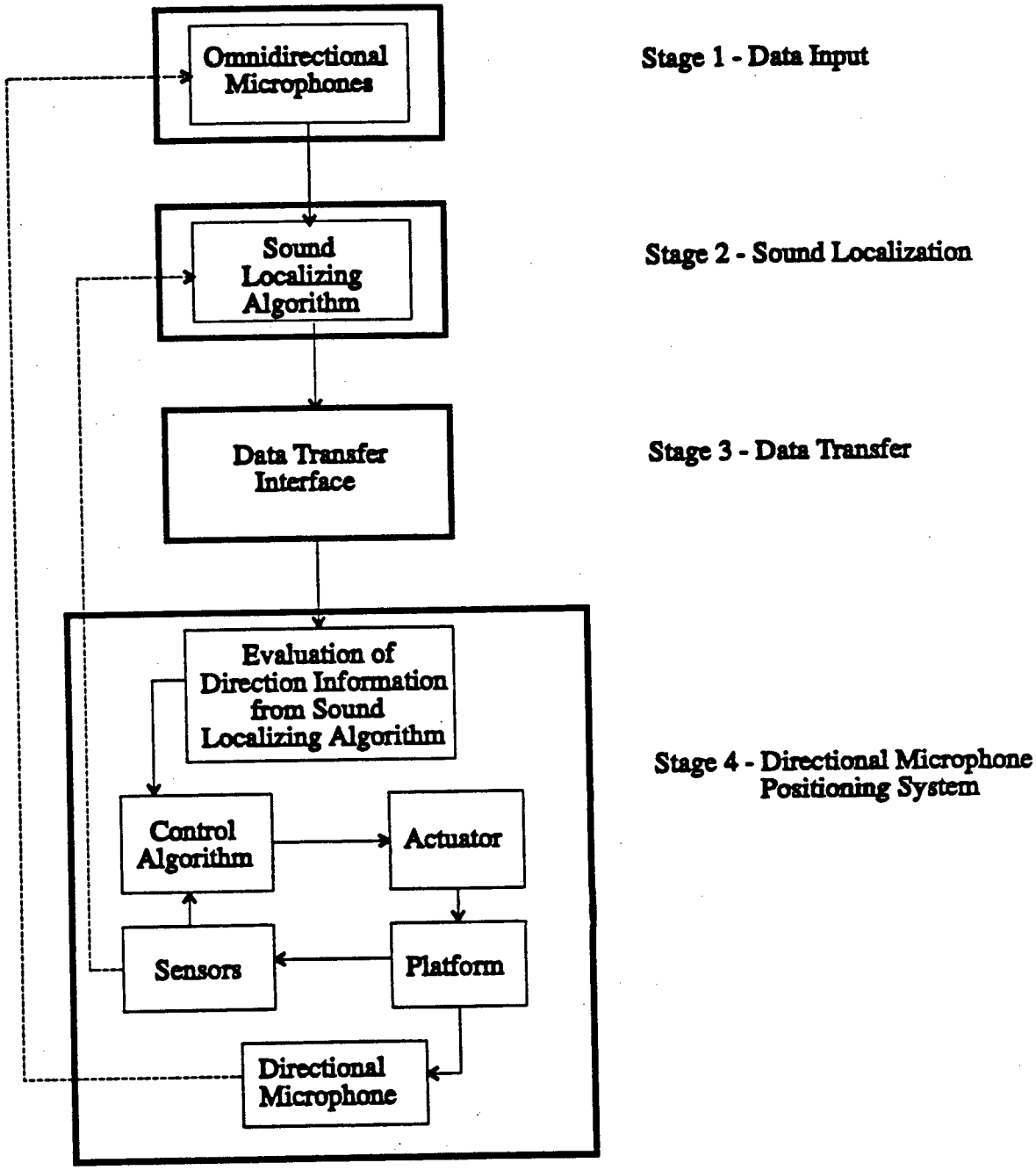


Figure 1: Information flow diagram of the conceptual design of the assistive listening device. Solid lines indicate the probable flow of information. Dashed lines indicate possible relationships between various components in the device, as considered in the design process.

from the sound localizing algorithm to the positioning system is defined as a stage in order to emphasize the independence of the sound localizing and directional microphone positioning stages. The first stage is the data input, consisting of omnidirectional microphones. The second stage is the sound localizing component. For this we require the data collected with the omnidirectional microphones, and a means of processing the collected samples to determine the location of any source present. The third stage is data transfer. This is the transfer of the results of the sound localizing algorithm to the directional microphone positioning system (DMPS). The fourth stage is the DMPS. This stage is further divided into the basic components likely to be required in the mechanical positioning of a directional microphone. The results of the sound localization process must be evaluated to determine the appropriate position for the directional microphone. A control algorithm is required to calculate the necessary control signal to apply to the motor. Position sensors and a feedback loop are required to monitor the position of the motorized platform on which a directional microphone would be mounted. In Figure 1, solid lines show necessary connections between components, and dashed lines indicate possible relationships between components in the device, which will be discussed later in the thesis.

A search of the literature and patent applications for assistive listening devices has not identified any device that combines sound localization based on human auditory function with a mechanically steered microphone to create a personal assistive listening device. The utilization of such a combination is the new approach that we take in this project.

1.4 Motivation, Project Objectives, and Scope

The objective of this work is to develop and test a prototype of a new assistive listening device. Specifically, we aim to develop a device that takes in data from a small array of omnidirectional microphones, uses the data to determine the location of any sound source, and

automatically steers a highly directional microphone towards the identified sound source. The output from the directional microphone should be suitable for direct input to a hearing aid. A device of this sort would be useful in controlled environments such as in the home, the classroom at a business meeting or some other situation where those involved are seated, there is tabletop space available for the device to sit, and there is orderly conversation where in general only one person is talking at a time. For the purposes of easy reference throughout this script, this device will be referred to as the **A**utomated **S**ound **L**ocalizing and **A**mplification **D**evice (ASLAD).

In addition to its use as a tool for the hearing impaired, there may be other applications for a device of this type. For instance, it could be used as the receiver for a teleconferencing system to improve sound quality for conference calls. It could also be used to provide input to a speech recognition device. This combination may be useful in the development of a portable and automated text interpreting system for the hard of hearing. Such an interpreter would enable a hard of hearing person to function in a business environment without the assistance of a typist who produces speech-to-text interpretation.

The goal of this project is to develop an adaptable system that would use an existing sound localizing algorithm to produce the positioning information for the directional microphone. The development of new sound localizing algorithms is beyond the scope of this work. However, the ASLAD design should be able to accommodate other sound localizing algorithms in future generations of the device.

1.5 Thesis Outline

The remainder of this thesis is divided into four chapters. Chapter two is devoted to the different topics that provide the theoretical background of this project. Chapter two begins with a summary of binaural hearing research and the properties of speech. These sections provide a

basis for the development of the sound localizing algorithm. A review of some signal processing techniques used in the implementation of the sound localizing algorithm follows. Finally, there is a summary of some feedback control methods, and definitions of parameters used to characterize the control of second order systems. These are used in the design and development of the DMPS. Chapter three contains a detailed description of the design and construction of ASLAD. Chapter four details the results of testing done, both on the performance of the individual stages of ASLAD, as well as on ASLAD as a whole. The tests of ASLAD as a whole were performed in an anechoic environment. Chapter five contains a discussion on the results of the tests and on various aspects of the design, and a summary of the work presented in this thesis. This concluding chapter also offers some recommendations for future work in this field.

CHAPTER 2: Theory

2.1 Introduction

In this chapter, the necessary background theory for the design of ASLAD is presented. In many instances during the design process, multiple ways to achieve the same goal were considered, each with varying levels of complexity. Our approach was to adopt a simple design philosophy. Wherever possible, simple, but adaptable, methods for accomplishing each of the necessary stages shown in Figure 1 were selected. Thus, only that information considered to be useful and relevant to the design as presented in chapter three is included in this chapter.

Section 2.2 provides a brief overview of some central concepts in the modelling of the human auditory system. This, combined with section 2.3 on the spectral and temporal characteristics of speech, provides the basis for the development of the sound localizing algorithm selected for use in ASLAD. The sound localizing algorithm is presented in section 2.4, and some signal processing techniques used in the implementation of the sound localizing algorithm are summarized in section 2.5. Section 2.6 is a summary of some relevant information on feedback control of mechanical systems, used in the design and control of the directional microphone positioning system (DMPS).

2.2 Binaural Hearing

Humans not suffering from any type of hearing loss or impairment have very sophisticated abilities to perform signal processing and interpretation of the sounds presented to them. We are capable of reverberation suppression, separating original sounds from echoes, selective listening (the ability to focus in on a desired sound source in an environment containing multiple sources), we can even follow two people speaking at the same time. In an anechoic chamber, humans can

identify the location of single sound sources to within one degree horizontally and five degrees vertically if the source is in front of the subject (Mills, 1958), and to within five degrees horizontally if the source is to one side of the subject (Hausler *et al.*, 1983). Three basic approaches have been taken to understanding and modelling binaural hearing. One approach is based entirely on psychophysical data. Such data is collected from experiments performed on humans, in which the ability to localize sound sources effectively under different conditions is assessed. These sorts of experiments are either carried out in the free acoustical field in a controlled environment such as an anechoic chamber, or are performed using headphones. With the headphone method, slightly different signals can be presented to either ear, representing either time delays or intensity differences between the two ears. This allows the effect of different factors on the perceived location of the sound to be studied methodically.

Another approach to modelling binaural hearing is based on physiological data. This involves modelling the response of neurons to sound stimulus, and developing a transfer function for the ear and the neurons activated by sound stimuli. Understandably, the data collection on neural activity is generally done on animals.

Finally, there is a collection of models based on a combination of psychophysical and physiological evidence.

There are a number of concepts that recur in many of the models, particularly in the psychophysical models. Looking at a diagram of the human head, we can see that the signal received at each ear from a single sound source can differ in three ways (Figure 2). Due to the difference in path length, there will be a temporal difference, a phase difference for each frequency, and an intensity difference. There may also be an intensity difference due to the shadowing effect of the head. These differences give us three different cues that should be examined to determine their usefulness in an automated sound localizer. They are commonly

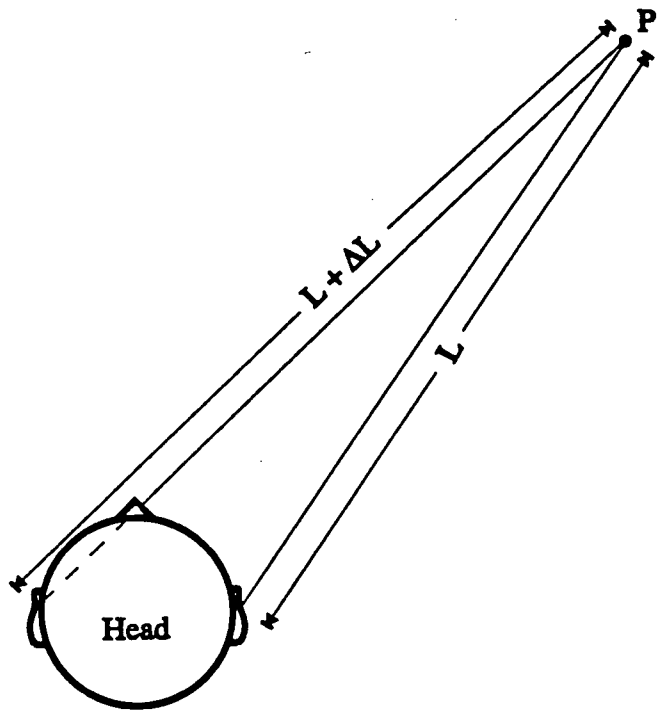


Figure 2: A difference in the path length of a sound originating at point P results in time, phase, and intensity differences between the signal received at the two ears. The shadowing effect of the head may cause an additional intensity difference between the two ears.

referred to in the literature as: interaural intensity difference (IID); interaural time delay (ITD); and interaural phase delay (IPD). The IPD is an equivalent to the ITD for continuous tones. Evidence that these cues are contributing factors to sound localization has been available for a number of years (Jeffress, 1948; Butler, 1969; Searle *et al.*, 1976).

It appears that the ear and the brain utilizes the IID mostly at high frequencies, and the ITD at low frequencies (below 1500 Hz). At low frequencies, the wavelength of sound is long. At higher frequencies, the wavelength is much shorter, and the head is more effective in shadowing the ear on the side further from the sound source. Also at higher frequencies, the wavelength of a continuous tone can be smaller than the inter-ear distance. If the signal is out of phase by more than 360°, then the IPD becomes ambiguous. Psychophysical studies investigating this problem have resulted in the proposal of a precedence effect. According to this theory, it is the first wavefront that is primarily used by the ear in the analysis of incoming signals. Psychophysical experiments using direct sounds and their echoes support the theory of a precedence effect. These studies have shown that by varying the time delay between the presentation of a direct sound and its echo, three different results can be obtained. Short delays in the presentation of direct and echo components result in a perceived direction intermediate to the directions of the two sounds. Delays of an intermediate length between the signals result in the earlier (direct) sound being localized. This intermediate region is called the precedence region. Long delays between presentation of the direct and echo signal result in the perception of two separate sounds incident from two separate directions. The delay at which two separate sounds are perceived is called the "echo threshold" (Colburn, in press). The use of echoes in precedence effect studies is also interesting because it is a special case of multiple source localization where the echoes are specifically related to the original source.

The way different cues are used under different circumstances is thought to depend on

the nature of the signal, *a priori* information about the signal and the environment, and the direction of incidence of the sound source. Furthermore, in the case of experiments, the subjects' knowledge of details about the experiment, and amount of "practice" with the experiment are also relevant. In real life situations, a person's ability to perform may be affected by modulation of the received signals by moving the head, reverberation patterns, and familiarity with the received signals (Colburn, in press). In the case of ITD and IID, the weighting of each factor in the localization process for different situations is often called time-intensity trading.

2.3 Properties of Speech

A brief summary of some of the basic characteristics of speech is presented here. There are two reasons for doing so. One is that some of these characteristics were taken into account in the design of the sound localizing algorithm. The other is that, although not implemented in ASLAD, a couple of simple methods have been studied to discriminate between speech and non-speech signals for assistive listening devices (eg. Flanagan *et al.*, 1985). These are based on matching the characteristics of the incoming signal with a defined set of characteristics for speech.

Fundamental frequencies in speech range from about 60 to 500 Hz, but an individual speaker will not normally use more than about an octave. Men produce the lowest fundamental frequencies, followed by women, then children. These frequencies are, on average, 120 Hz for men, 225 Hz for women, and 265 Hz for children, but during speech they vary continuously about these mean values (Fry, 1979). An example of the Fourier transform of a speech sample is shown in Figure 3(a). The spectral characteristics are basically low pass in nature, falling off at about 8-10 dB/octave above approximately 500 Hz (Fry, 1979). Most of the energy in speech is concentrated in the range from dc to 1 kHz. The fundamental frequency as well as a number

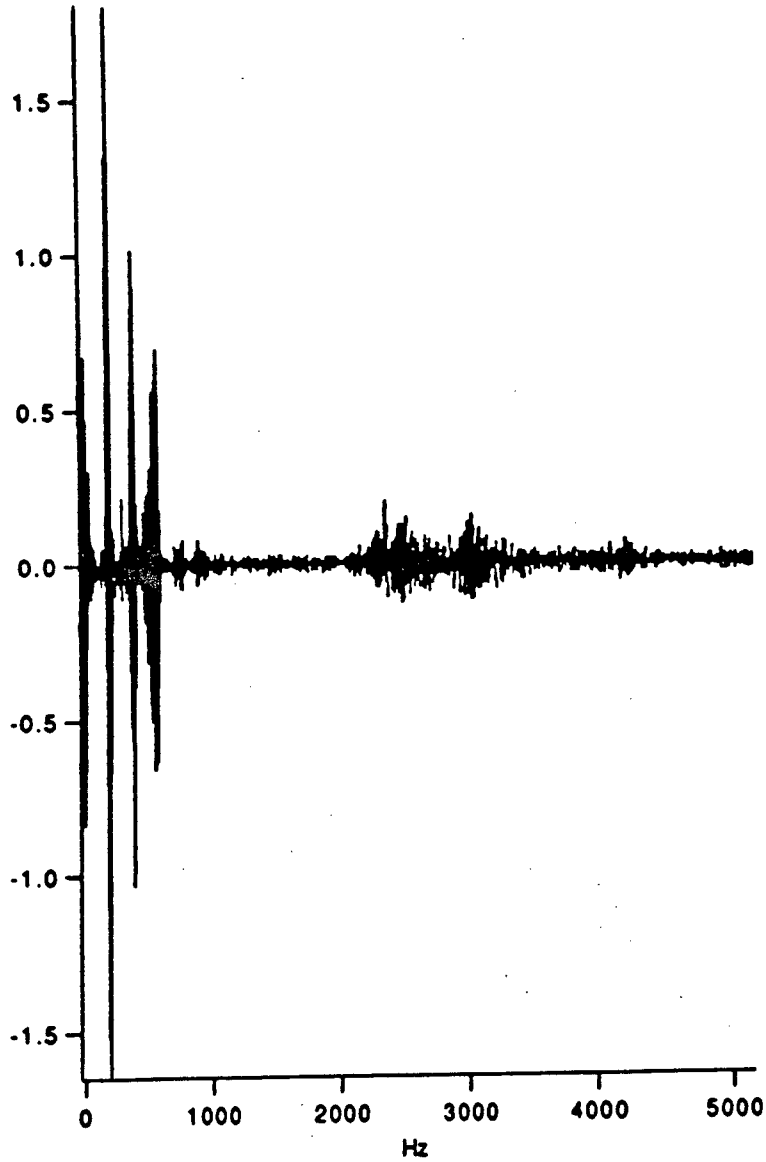
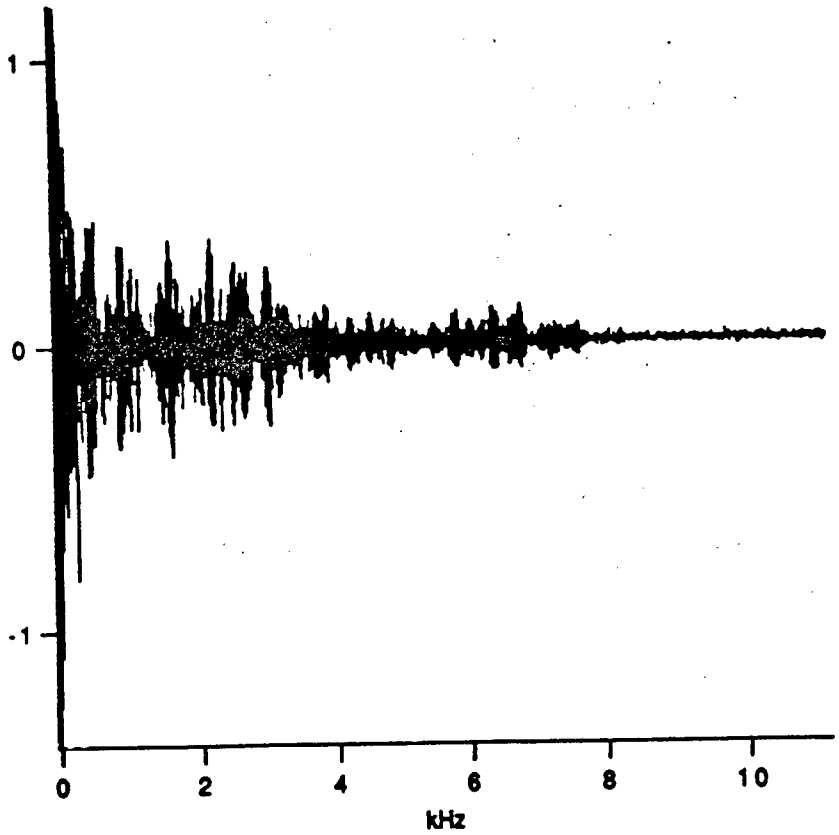
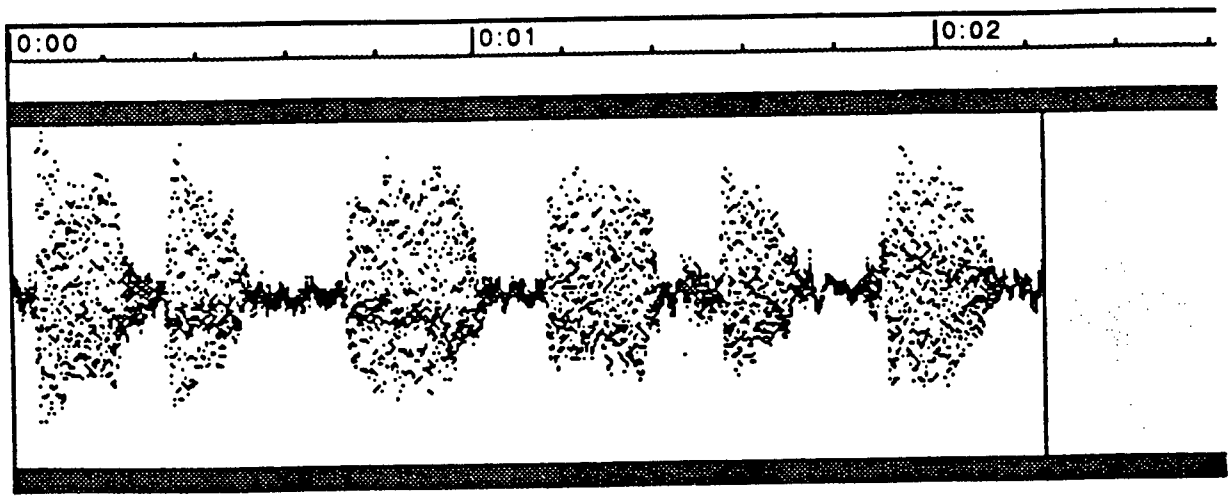


Figure 3: (a) Fourier Transform of a speech sample, showing the relative amplitude of the fundamental and secondary harmonics, and the typical frequency ranges in which the energy in speech lies; (b) Fourier Transform of a man-made noise (relative amplitude vs. frequency); (c) time series for the phrase "a-b-c-one-two-three" (relative amplitude vs. time).


1
0
-1
0
2
4
6
8
10
kHz


0:00
0:01
0:02

of higher order harmonics are clearly visible in the spectrum. There are generally some other resonant frequencies, as can be seen in Figure 3(a), in the range from 2 to 4 kHz. These can be attributed to the airways leading from the larynx (location of the vocal chords) to the mouth and nose. In contrast, the spectrum of many common noises will not exhibit the pattern of a fundamental frequency with higher harmonics, and will have a more uniform distribution of energy over the frequency range from dc to 5 kHz. Figure 3(b) is an example of a man-made noise.

One other important quality of speech is that it tends to be "bursty" in time. Figure 3(c) is the time series (intensity plotted as a function of time) for a speech sample of the phrase "a-b-c-one-two-three". The envelope of this, or any other speech sample will give periodic peaks in amplitude. The point of largest amplitude is commonly at the beginning of a word or syllable. This is a result of the way we form different sounds by motions such as forcing the tongue up against the roof of the mouth and then releasing it, causing a burst of air to flow out.

The latter property, based in the time domain, is the basis of the sound localizing algorithm used in ASLAD.

2.4 Sound Localizing Algorithm

There are many different versions of sound localizing algorithms that could have been implemented in this prototype. In fact, those studying sound localization are constantly modifying and improving models of binaural sound localization. However, not all of the cues are appropriate for use in ASLAD. The IID is not easily obtainable because the intensity difference between sensing microphones separated by only 20 cm would be insignificant without the shadowing effect of the human head. Even if there was a measurable difference, accuracy would demand that the gain of sensing microphones be matched very precisely, and the

amplifying circuitry be stabilized to compensate for component drift. Otherwise, slight changes in ambient temperature could result in large errors in localization. The IPD cue is frequency dependent, and would require comparisons between signals for a large range of frequencies. There are also complications due to multiples of 2π differences in phase for frequency components whose wavelength is shorter than the separation of the sensing microphones. The ITD cue is frequency independent, and it is used by humans predominantly for lower frequency input. Much of the energy in speech is concentrated at the low end of the spectrum, coinciding with the range at which the ITD cue dominates in binaural sound localization. At present, this is the most practical cue on which to base a sound localizing algorithm for ASLAD. An algorithm based on this cue will be possible to implement in near real time.

The sound localizing algorithm selected for this project was developed by Zakarauskas (patent application, 1993), and is similar to that found in (Middlebrooks and Green, 1990). The process is described below:

(1) For each plane in three-dimensional space signals from two omnidirectional microphones separated by some distance.

(2) The signal from each omnidirectional microphone is rectified.

(3) The signal from each omnidirectional microphone is passed through a low-pass filter to obtain the envelope of the signal.

(4) The local maxima of each signal are identified and replaced with impulse functions. The relative amplitude of each maximum is maintained in this step.

(5) A cross-correlation of the two arrays of impulse functions is performed. This yields a single peak for each sound source. For a sound source located directly in front of, and equidistant from, the microphones (see A in Figure 4(a)), the cross-correlation will yield

a peak at zero. If a given sound source is located elsewhere (see **B** in Figure 4(a)), the time delay between the signal's arrival at each microphone is represented by a shift in the peak of the cross-correlation function. This shift is a measure of the angle by which the source is "off centre".

Figure 4(a) is a diagram showing a pair of omnidirectional microphones. For a sound source originating at a point **B**, the angle of incidence $\phi_{1\text{-}2}$ is determined using a plane wave approximation. It is given as:

$$d\cos\theta = c \times ITD$$

If the signal is sampled at evenly spaced intervals (denoted by the symbol Δ), then

$$ITD = n \times \Delta$$

$$\therefore\ d\cos\theta = cn\Delta$$

$$\cos\theta = \frac{cn\Delta}{d}$$

$$\theta = \arccos\left(\frac{cn\Delta}{d}\right)$$

$$\therefore\ \phi_{1-2} = 90° - \theta = \arcsin\left(\frac{cn\Delta}{d}\right)\ \ degrees \qquad (2.1)$$

where c is the speed of sound in air, n is the bin number representing the number of sampling intervals between arrival of the signal at the first and second microphone (obtained by cross-correlation), Δ is the sampling interval, and d is the distance between the two microphones. At least one other system has been shown to give reliable results using the plane wave approximation even at the relatively close distance of one meter (Flanagan *et al.*, 1985).

Since the data is digitally sampled, the number of different angles that can be selected is finite. For example, if the distance between the sensing microphones is 20 cm, and the sampling

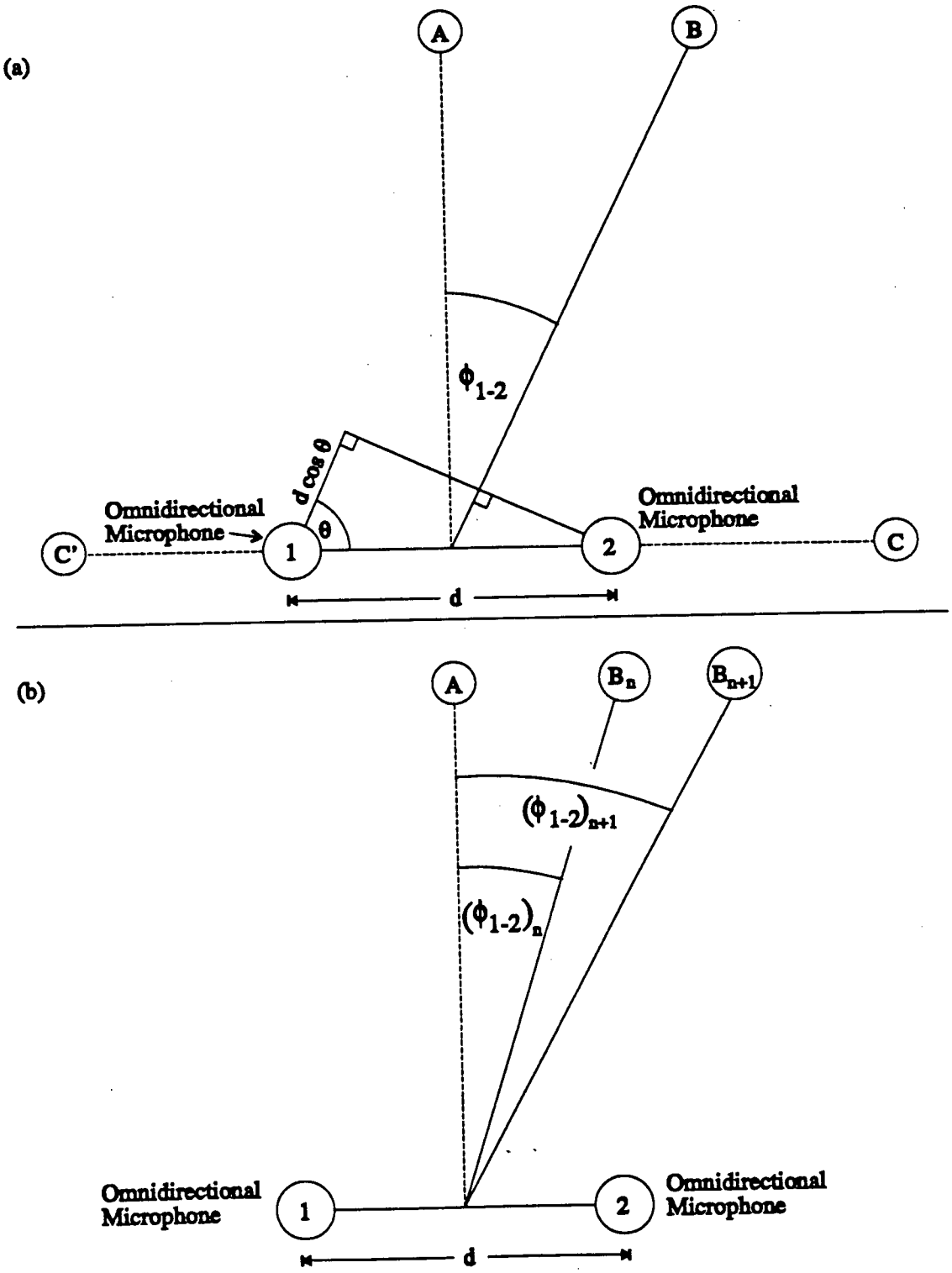
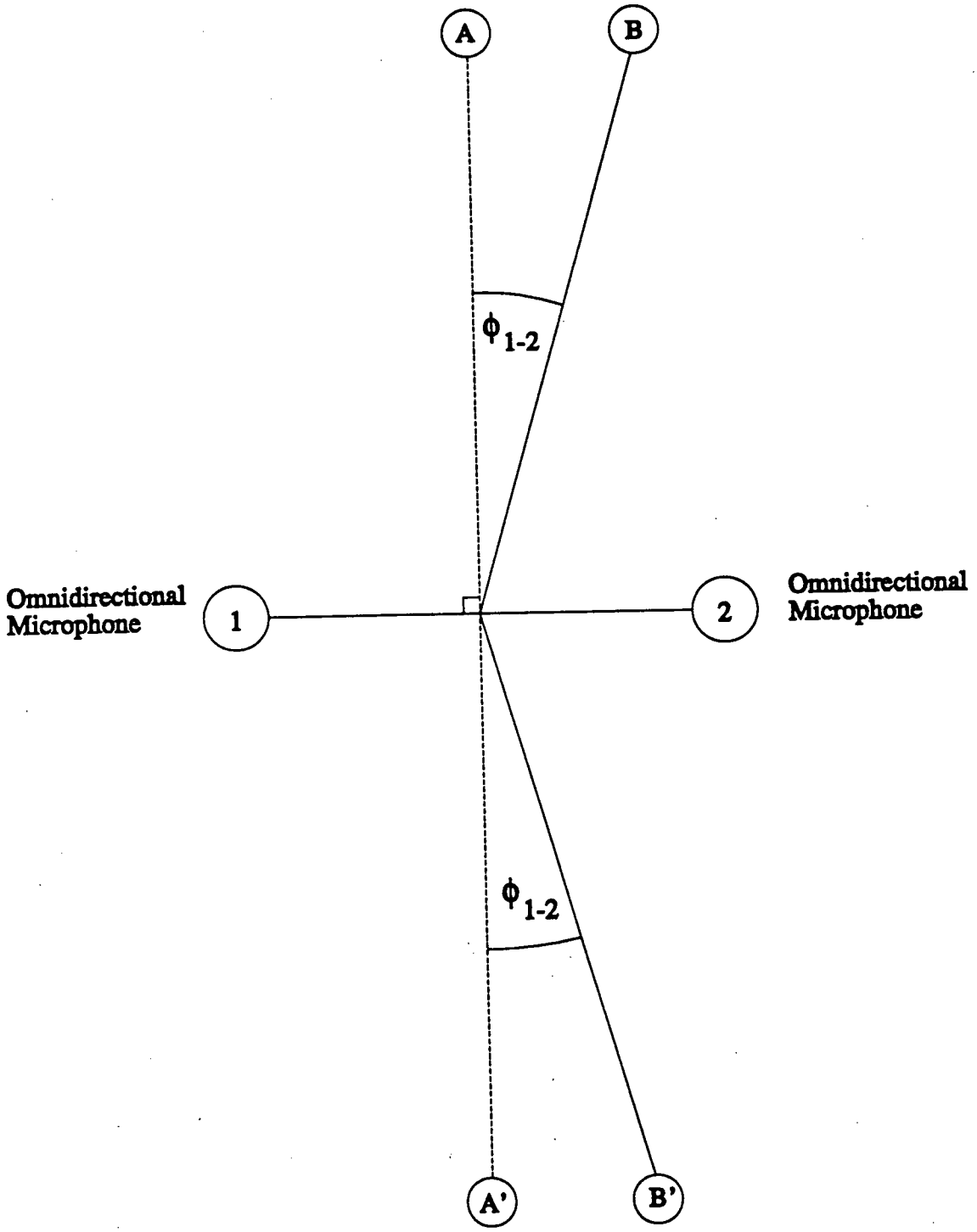


Figure 4: (a) Geometry for the derivation of equation (2.1), angle of incidence (ϕ_{1-2}) of a target at a position **B**, for a pair of omnidirectional microphones; (b) geometry for the derivation of equation (2.2), the localization resolution for the angle derived in equation (2.1); (c) for each pair of omnidirectional microphones, there are two possible target positions for each solution to the sound localizing algorithm (eg. **B** and **B'** for the angle ϕ_{1-2}).

(c)


A
B
ϕ_{1-2}
Omnidirectional Microphone
1
2
Omnidirectional Microphone
ϕ_{1-2}
A'
B'

rate is 20 kHz (for which the sampling interval is 0.05 milliseconds), there will be 23 *bins*, where each bin represents one sampling interval, into which the solution could fall. For a source at position **C** in Figure 4(a), there will be a maximum time difference between signal arrival at each microphone of 11 sampling intervals. For a similar source located in the opposite direction (see **C'** in Figure 4(a)), there will also be a maximum time difference of 11 sampling intervals. However, in the cross-correlation, one of these differences will show up as a lead (peak at n = 11) and the other as a lag (peak at n = -11). For a source equidistant from the two microphones (**A** in Figure 4(a)), there will be zero time difference, and the cross-correlation result will be a peak at n = 0.

The resolution due to the digitization of the data is dependent upon the sampling rate and the angular position of the source. Figure 4(b) is a diagram showing the relationship between position and error. Say points $\mathbf{B_n}$ and $\mathbf{B_{n+1}}$ on Figure 4(b), at angles of $(\phi_{1\text{-}2})_n$ and $(\phi_{1\text{-}2})_{n+1}$ respectively, represent positions corresponding to the solution of equation (2.1) for two consecutive values of n. Any source originating from an angle between $(\phi_{1\text{-}2})_n$ and $(\phi_{1\text{-}2})_{n+1}$ can at best be localized at either $(\phi_{1\text{-}2})_n$ or $(\phi_{1\text{-}2})_{n+1}$. The resolution at a particular angle is estimated as the angular separation between two consecutive solutions to equation (2.1). It can be written as a function of $(\phi_{1\text{-}2})_n$:

$$resolution((\phi_{1-2})_n) = \pm\left(\arcsin\left(\frac{nc\Delta}{d}\right) - \arcsin\left(\frac{(n+1)c\Delta}{d}\right)\right) \tag{2.2}$$

The resolution as a function of the angle $\phi_{1\text{-}2}$ is plotted in Figure 5 for a number of different sampling rates. From this graph it can be seen that for a given sampling rate, the sound localizing algorithm is most accurate at small values of $\phi_{1\text{-}2}$. At locations where $\phi_{1\text{-}2}$ is large, there is a large error.

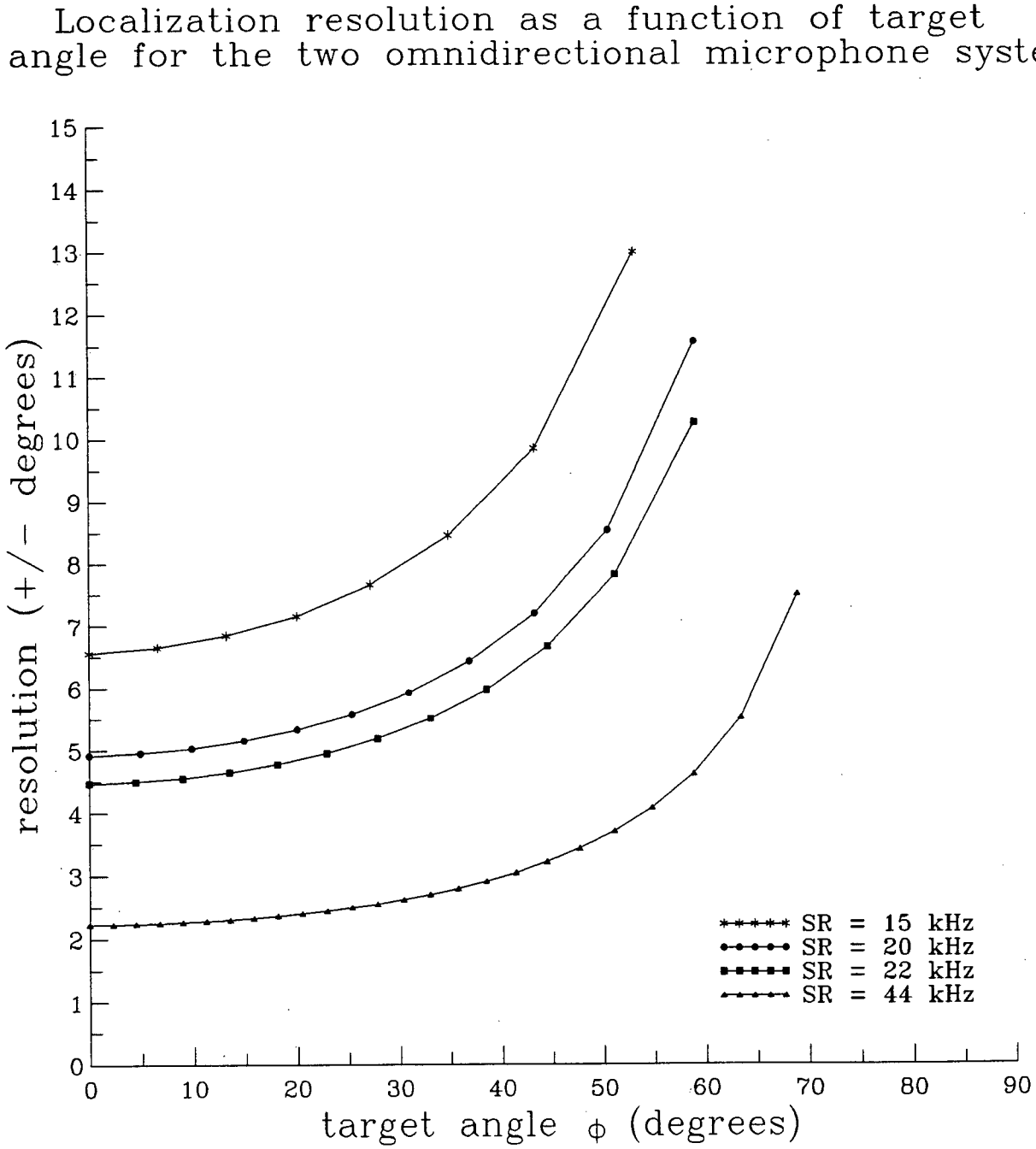


Figure 5: Localization resolution as a function of target angle (ϕ) for a number of sampling rates (SR).

It must be noted that the cross-correlation function does not yield a unique solution. Figure 4(c) shows the same microphone pair as before. Also shown are two sources, **B** and **B'**. For every source **B** in the front half plane of two sensing microphones, there is an image source **B'** in the back half plane that would give the same solution to the cross-correlation function. This problem was considered in the design of ASLAD, and will be discussed in section 3.3.1.

2.5 Signal Processing

2.5.1 Fourier Transforms

The use of Fourier transforms in signal processing is common, and the theory and methods can be found in almost any general text on communications or signal processing. Those used to develop this summary were (Haykin, 1989; Proakis and Manolakis, 1992), and for information on practical applications, (Press *et al.*, 1986).

It is often desired to study the frequency components of a signal. There are many different ways to obtain the frequency domain equivalent of a signal. Examples include: cosine transforms, Hartley transforms, wavelet transforms, and Fourier transforms. The Fourier transform is a popular and well developed method for converting a signal described in the time domain to its frequency domain equivalent, and was thus chosen for use in programming the sound localizing algorithm. Even if the frequency components of a signal are not of interest, Fourier transform methods can be used to perform the cross-correlation of two signals more efficiently than it could be done in the time domain. In the design presented in this thesis the Fourier transform methods are used to perform the functions of filtering and cross- correlation, as required for the sound localizing algorithm, but the information about the spectrum of the collected signals is not used.

For a continuous function, the Fourier transform is defined as:

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{2\pi i f t} dt \tag{2.3}$$

and the inverse Fourier transform is given by:

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{-2\pi i f t} df \tag{2.4}$$

The Fourier transform method can also be used for data collected at evenly spaced intervals in time (Δ). This is called a discrete Fourier transform, and it is derived by approximating the integral in equation 2.3 by a discrete sum. The transform is given by:

$$H_n \equiv \sum_{k=0}^{N-1} h_k e^{2\pi i k \frac{n}{N}} \tag{2.5}$$

where N is the number of data samples, and n = -N/2,...N/2. This transform maps N complex numbers in the time domain (the h_k's) into N complex numbers in the frequency domain (the H_n's). The value f_n is the frequency value of the nth data point in the transformed series, and it is given by $f_n = n/N\Delta$. If the set of numbers being transformed are the samples of a continuous function sampled at the interval (Δ) then relationship between the discrete and continuous transforms is approximated by $H(f_n) \approx \Delta H_n$. The inverse discrete Fourier transform is given by:

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n e^{-2\pi i k \frac{n}{N}} \tag{2.6}$$

The sampling rate is given by the inverse of the time interval between samples collected, and must be at least two times the highest frequency component in the signal being collected in order to be able to reproduce the signal and to avoid aliasing. This value is called the Nyquist critical frequency ($f_c = 1/2\Delta$). *Aliasing* occurs when frequency components outside of the range ($-f_c$,f_c) are falsely translated into that range. If the sampling rate cannot be made high enough

such that f_c is twice the natural frequency limit of the signal being sampled, a low-pass filter must be applied at the input stage, before the digitization process, so that no unwanted components of the signal are aliased.

The Fourier transform is computationally time consuming, but an algorithm called the Fast Fourier transform (FFT) can be used to reduce the computation time for the discrete Fourier transform. The original Fourier transform requires a number of computations of the order N^2 while the Fast Fourier transform only requires computational steps on the order $N\log_2 N$ (where N is the number of points to be transformed). For this algorithm, the length of the series must be a multiple of two. The algorithm uses bit reversal to rearrange the order of the data in the array, and breaks down the array into transforms of length 2, 4, 8, up to N. The basic operation in the FFT is the swapping of two data points. The FFT can be performed such that the transformed result is stored in the same memory block as the original array, which reduces the memory requirements.

In the general case, a signal is described by a complex array of data for which the transform is also complex. The representation of each complex data point requires two spaces in the array: one for the real, and one for the imaginary component. The transform of a set of data that is real in the time domain is generally complex, but the transformed data can be packed into half of the required space by taking advantage of symmetries. This allows one to perform the transform of a real array of data using the same amount of space as for a complex array of half the length. This is another method for reducing the memory requirements for the DSP chip used. It is used in the algorithms implemented in ASLAD.

2.5.2 Convolution and Digital Filtering

The Fast Fourier transform can be used to perform some signal processing operations

more efficiently than they could be done in the time domain. Two such operations are cross-correlation and convolution. Convolution of two signals in the time domain is described by equation 2.7:

$$a * b \equiv \int_{-\infty}^{\infty} a(\tau) b(t-\tau) d\tau \Leftrightarrow A(f)B(f) \tag{2.7}$$

The Fourier transform of convolution is the multiplication of the individual Fourier transforms of the two signals in the frequency domain. The equivalent of equation 2.7 in a discrete form is:

$$(r * s)_j \equiv \sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} s_{j-k} r_k \Leftrightarrow S_n R_n \tag{2.8}$$

Convolution in the time domain is the equivalent of a digital filter in the frequency domain. Once the array of data is transformed into the frequency domain, it can be multiplied by the filter transfer function, and transformed back into the time domain. In order to get mathematically correct results, there are some extra steps necessary. The filter function must be even and real if the filtered data is to be real. Before being transformed, the end of the data array must be padded with zeros so as not to affect the data in a manner similar to the aliasing described above. For a symmetrical filter function, the number of zeros in the padding must be equal to half the length of the filter function (in the time domain) in order not to corrupt the data at either end of the array. The shape of the filter transfer function is also important. The use of a step function at the desired cutoff frequency results in the addition of oscillations in the time domain signal. This is known as *ringing*. The filter transfer function should be quite smooth in order to avoid ringing in the filtered data.

2.5.3 Cross-Correlation

Cross-correlation is another operation that is more efficiently computed in the frequency domain. The cross-correlation of two continuous functions is given by equation 2.9, and its transform is equivalent to the Fourier transform of one function multiplied by the complex conjugate of the Fourier transform of the other:

$$Corr(a,b) \equiv \int_{-\infty}^{\infty} a(\tau + t)\, b(\tau)\, d\tau \Leftrightarrow A(f)B^*(f) \tag{2.9}$$

In discrete form, this is written as:

$$Corr(r,s)_j \equiv \sum_{k=0}^{N-1} r_{j+k} s_k \Leftrightarrow R_k S_k^* \tag{2.10}$$

This assumes a periodic function, which, in general, will not be the case. To avoid any errors in the cross-correlation of non-periodic functions due to end effects, the ends of the arrays being correlated must be padded with at least as many zeros as the size of the maximum lead or lag possible in the system. The correlation of two arrays of data is performed by taking the FFT of each set of data, multiplying one by the complex conjugate of the other and then performing the inverse transform on the resulting data set in order to return to the time domain. The cross-correlation of two real arrays of data will also be real. For two identical signals, the results of the cross-correlation will be a peak value at the n=0 element of the solution array. For two signals identical in form but offset in time, the results will be a peak value at a position n of the solution array. The value of n times the sampling interval gives the time delay or offset between the two signals, and the sign of n, positive or negative, represents either a lead or lag of one signal relative to the other. For two unrelated signals, such as random noise, there will generally be no peak in the cross-correlation solution array.

2.6 Control Theory

It is possible to control a mechanical system such as the proposed DMPS with great accuracy (eg. error in positioning of 1° or lower). However, for ASLAD this may not be necessary, and may not, in fact, be desired. High accuracy positioning will generally have a cost of one type or another associated with it, such as expensive precision sensors and mechanical components, or processor-intensive control algorithms. In this section, some information about the characteristics of typical motor and load systems, and of basic feedback control techniques are presented. This information was used in both the design of the mechanical components of ASLAD, and also in the control algorithm for the DMPS.

2.6.1 The Mechanical System

In Figure 1, the basic components required in a position control system were shown. They consisted of an actuator (a motor), a platform on which a directional microphone could be mounted, and one or more sensors to provide information about the position and/or motion of the directional microphone. The results from the sound localizing algorithm and the information from the position sensors would be the sources of input for the control algorithm. The directional microphone and platform can be grouped together and referred to as the load.

For a typical motor and load system as shown in Figure 6, the total voltage applied to the armature (ε_a) is given by the differential equation:

$$\varepsilon_a = L_a \frac{di_a}{dt} + R_a i_a + \varepsilon_b \tag{2.11}$$

where L_a and R_a are the inductance and resistance of the armature circuit respectively. The torque (τ) produced by applying the armature current is used to overcome inertial and frictional

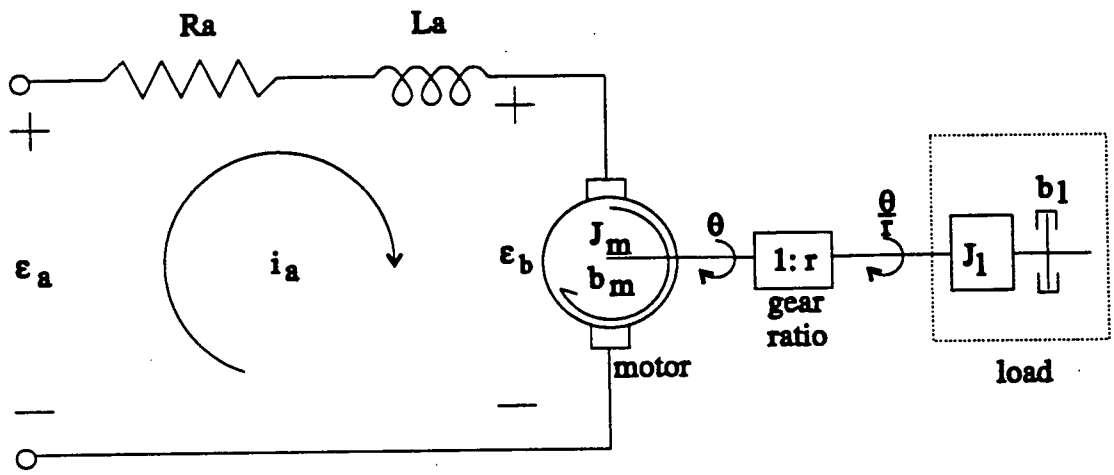


Figure 6: Basic circuit diagram for a motor and load.

forces in the system being driven, which are related to the motion of the motor shaft as follows:

$$J_{eq}\frac{d^2\theta}{dt^2} + b_{eq}\frac{d\theta}{dt} = \tau = Ki_a \tag{2.12}$$

where J_{eq} is the equivalent inertia and b_{eq} is the equivalent coefficient of friction, for which definitions will be given below. K is a motor-torque constant.

There are two different sources of inertia and friction that need to be considered in the rotating platform system: one is the motor, and the other is the load (platform and directional microphone). The inertia and friction associated with the load can be referred back to the motor shaft as follows, resulting in the equivalent terms:

$$J_{eq} = J_m + r^2 J_l \qquad (r<1) \tag{2.13(a)}$$

$$b_{eq} = b_m + r^2 b_l \qquad (r<1) \tag{2.13(b)}$$

where J_m is the inertia of the motor, J_l is the inertia of the load, b_m is the coefficient of friction associated with the motor, b_l is the coefficient of friction associated with the load, and r is the ratio of the gear on the motor axis to the gear on the load axis (Ogata, 1990). As can be seen from the equations, if the gear ratio is small enough then the effective additional inertia and friction at the motor shaft is negligible.

The motorized platform as described above is a basic second order system. Second order systems are very common in engineering practice, and the set of parameters used to describe second order systems is defined in most texts on control (eg. Franklin *et al.*, 1986; Ogata, 1990) These parameters will be described in section 2.6.3, as they are used to characterize the performance of ASLAD.

2.6.2 Feedback

A system becomes more robust with the addition of a feedback loop, as it can respond to unexpected disturbances or behaviour in its operation. For ASLAD, we are interested in the position of the directional microphone, and would like to be able to measure its position to ensure that it reaches the location determined by the sound localizing algorithm.

Position can be measured in two ways: directly, or indirectly. In the direct form, the feedback signal is a unique value representing one position. In the indirect form, a change in position is measured relative to some known starting point. In the latter, this measurement is often acquired from a shaft encoder mounted on the shaft of the motor or platform. This requires constant monitoring. In the direct form, an absolute position shaft encoder can be used in a single rotation system, or a potentiometric circuit can be used. The latter is used in a voltage divider circuit, which is calibrated so that there is a known relationship between the voltage measured across the potentiometer and position of the platform. One advantage of direct position sensing is that there will be no integration of the error over time, as can occur if the position is measured indirectly.

Three of the most common ways of using the feedback information are to develop control signals that are proportional to the feedback signal or its derivative, or to the integration of the feedback signal over a set period of time.

Proportional feedback is when the control signal is made proportional to the difference between the existing and desired value of the quantity being measured. In other words, the greater the error, the larger the magnitude of the control signal, resulting in a faster reduction of the error. This form of feedback signal is appropriate in position control.

For *derivative feedback*, the control signal is proportional to the rate of change of the error. Derivative feedback is generally used in conjunction with proportional feedback, integral

feedback, or both. It has the effect of increasing the damping and improving the stability of the system.

In *integral feedback* the control signal is proportional to the integrated, or accumulated, error over a given period of time. With this form of feedback, a control signal can be provided even when there is no error. The feedback signal is a function of past values of the error rather than the present one. With integral control, steady state error can be reduced and even eliminated, however, stability is often reduced as a result. This type of control scheme is most useful for some quantity that is to be maintained at a steady value for long periods of time, such as temperature.

For ASLAD, proportional feedback alone could be used, as could proportional feedback in conjunction with derivative feedback. Integral feedback is not appropriate in this type of position control system, as the frequent changes in the desired position of the directional microphone would result in the accumulation of large error terms. The directional microphone would never be able to settle at the desired position in the required time span.

2.6.3 Parameter Definitions

There are a number of standard parameters that are used to describe the performance of a control system. They are: delay time, rise time, peak time, maximum overshoot, and settling time. These parameters describe the transient response, usually measured by applying a step input to a system at rest. This section summarizes those used to describe a typical second order system. In Figure 7 is shown a generalized response marked with the parameters discussed below.

Delay time (t_d) is the time required for the response to reach 50% of the desired value. It is measured on the initial rise as shown in Figure 7. For ASLAD, we are interested in the

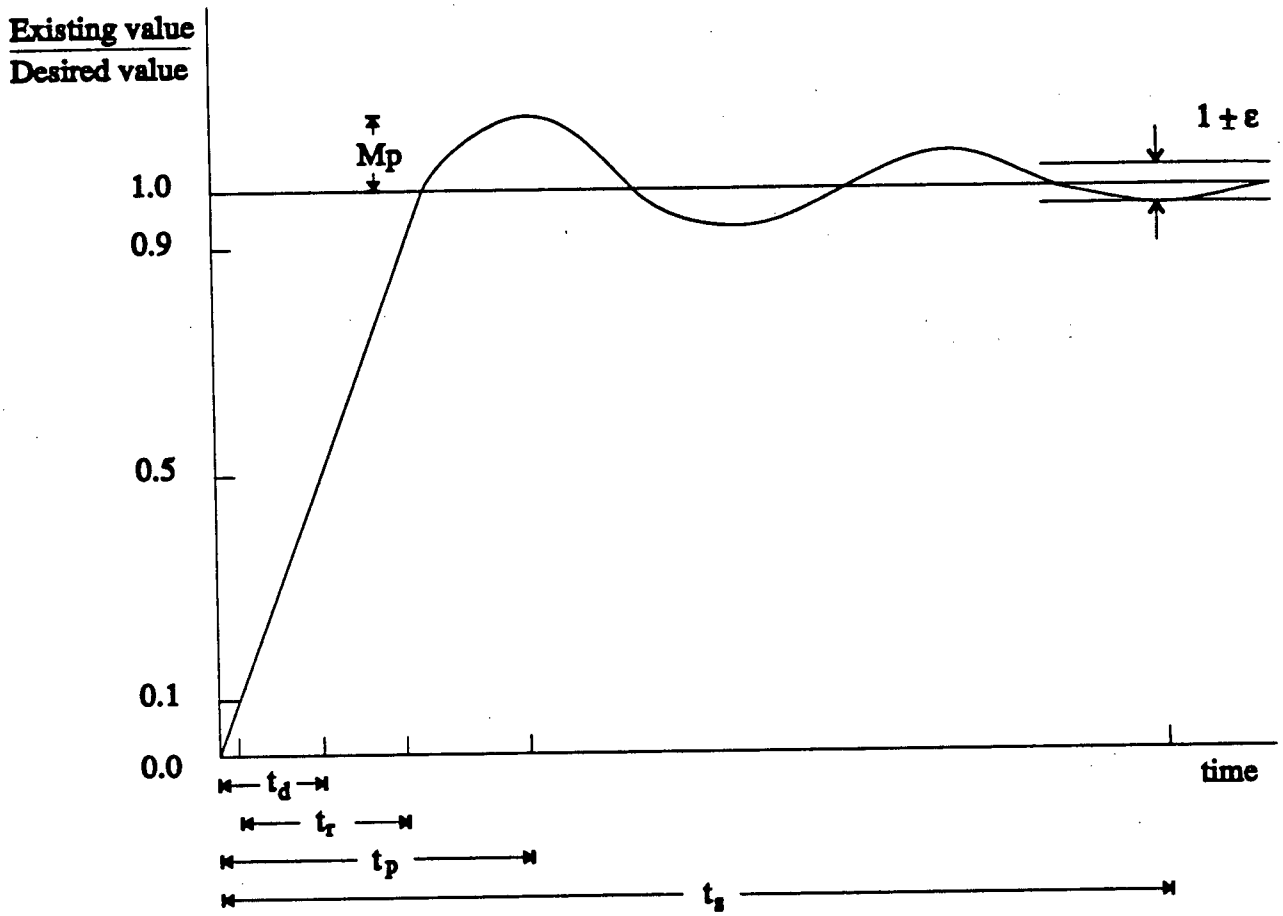


Figure 7: Definition of parameters used to characterize a typical second order system. Shown are delay time (t_d), rise time (t_r), peak time (t_p), maximum overshoot (M_p), and settling time (t_s).

delay time of the velocity response of the platform.

Rise time (t_r) is defined for underdamped systems as the time taken to go from 0% of the desired value to 100%. In the case of overdamped systems, the time taken to go from 10% to 90% of the final value is often used. For ASLAD, there are two rise times that could be measured, one for velocity response and one for position response. However, a measure of the time taken for the directional microphone to get to the desired position cannot, in general, be used as an indicator in this case, as the distance between the existing and desired positions is variable. Instead, a similar parameter has been defined, which is the maximum time allowed for travel between the two extreme positions of the rotating platform.

Peak time (t_p) is the time required for the response to reach the first peak of the overshoot. This parameter is not of much interest in the design of ASLAD. It is the magnitude of the overshoot that is of greater interest.

Maximum overshoot (M_p) is the amount by which the first peak of the response curve exceeds the final value, and it is generally measured as a percentage of the final, or steady state, value of the response. For ASLAD, we are interested in position overshoot. In the case of this design, where the desired position of the directional microphone is constantly changing, it is more appropriate to measure the maximum overshoot in absolute terms instead of as a percentage of the final value.

Settling time (t_s) is defined as the time required to reach the required output value within an acceptable range of error, and stay within that range. For ASLAD, settling time is used to describe the time taken for the motor and platform combination to reach travelling speed. This is the settling time for the velocity response.

2.6.4 Starting and Stopping

The most energy consuming part of the platform-control process is the start from a stationary position. This is because generally static frictional forces are larger than the frictional forces of a system in motion. In some mechanical systems a low magnitude ac ripple current is applied to a system at rest to keep it moving slowly. This is done so that static frictional forces do not need to be overcome when moving from one position to another. This is not an appropriate method for ASLAD, as the continuous motion of the directional microphone may prove distracting to those using it. Another method of overcoming static frictional forces and getting to operating speed as quickly as possible is to provide a current pulse to help a system overcome the frictional forces and start moving, and then scale back the driving current to a lesser value, as required to obtain the desired velocity. A current pulse of the opposite polarity can be used to bring a system to rest in the minimum amount of time.

CHAPTER 3: Implementation

3.1 Introduction

In this chapter, the design and construction of ASLAD is described. The initial stage of the design process was to decide exactly what would be built, and what performance specifications should be met. This chapter begins with an explanation of these design specifications. This is followed by a detailed description of the design and construction of ASLAD. The description is broken down into the categories of mechanical, electrical and software components.

3.2 Design Specifications

Table 3.1 is a list of the performance specifications that were chosen for ASLAD. In some cases there are ranges of acceptable values. Other values were dependent on the equipment used (eg. input impedance of the omnidirectional microphones), and were documented to facilitate any future changes. The table is divided into four different categories to reflect the different stages of the project as were shown in Figure 1. The specifications are discussed in the following four sections.

3.2.1 Stage 1 -- Data Input Design Specification

Stage one is data input. Four parameters have been defined for this stage: sampling rate, filter bandwidth, voltage range, and output impedance of the omnidirectional microphones.

It was shown in section 2.4 that the resolution of the sound localizing algorithm is dependant upon the *sampling rate* of the data collected from the omnidirectional microphones. Higher sampling rates give a better resolution. Therefore it is important to have a high sampling

Table 3.1: Design Specifications sheet for ASLAD.

STAGE	PARAMETER	PREFERRED VALUE	ACCEPTABLE LIMIT	COMMENTS
1. Input stage	Sampling rate	44.1 kHz		subject to processor capabilities
	Filter Bandwidth	10 kHz		
	Voltage range		±2.5 volts	range of A/D converter
	Output impedance of omnidirectional microphones			nominal value
	Omnidirectional microphone separation	20 cm		approximate separation of human ears
2. Sound Localizing Algorithm	Resolution			varies with sampling rate and target angle
	Processing time			
3. Data transfer	Time between updates	100 msec		subject to processor capabilities
	Data format			
4. Directional Microphone Positioning System (DMPS)	Total range (x) of system	270 degrees	180 degrees	
	Time allowed for x degrees of travel	1 second		
	Overshoot	±5 degrees	±10 degrees	
	Settling time	100 msec		must be ≤ time between updates
	Travelling speed	270 degrees/second	180 degrees/second	
	Tracking speed			not a design criterion
	Error on feedback signal	±2.5 degrees	±5 degrees	
	Mean positioning error			
	Max. positioning error			

rate. An industry standard for digital recording and reproduction of sound is 44.1 kHz, and so this was chosen as the initial value for the sampling rate.

The *filter bandwidth* had to be at least half of the sampling rate to satisfy the Nyquist criterion (and prevent aliasing) as noted in section 2.5.1. However, for speech, a much lower bandwidth will suffice, so a 10 kHz low-pass filter was chosen as a suitable value.

The *voltage range* for the input was bounded by the operating range of the analog-to-digital (A/D) converters used.

A record of the *output impedance* of the omnidirectional microphones was necessary for the design of amplifying circuitry, but a particular impedance value was not sought in the design of the input stage.

An *omnidirectional microphone separation* of 20 cm was selected. The larger the separation between the omnidirectional microphones, the higher the resolution achievable by the sound localizing algorithm. However, a personal assistive listening device must be small enough to be portable. This value was selected because it is approximately equivalent to the separation of human ears, and because the resulting size of the device would compare to a textbook or a laptop computer.

3.2.2 Stage 2 -- Sound Localizing Algorithm Design Specifications

Two parameters were defined for the sound localizing algorithm stage. They are resolution and the processing time.

The *resolution* is dependent upon the sampling rate used and position of a sound source relative to the pair of omnidirectional microphones. Because this parameter is so dependent upon other factors, an initial value was not chosen for it. However, high resolution of the sound

localizing algorithm is one of the keys to a successful device, and every attempt was made to ensure that the resolution was optimized in the design process.

The *processing time* required for completion of one cycle of the sound localizing algorithm could only be given upon completion of the software.

3.2.3 Stage 3 -- Data Transfer Design Specifications

The third stage is the transfer of data from the sound localizing algorithm to the DMPS. The two parameters for this stage are time between updates and data format.

The *time between updates* is the time required for one cycle of data acquisition, and execution of the sound localizing algorithm. A data acquisition cycle on the order of 100 milliseconds was chosen as the design value. Glottal pulses and other sharp vocal features, which give rise to the peaks in the time-intensity profile of a sound sample, are approximately 10 milliseconds in duration. In order to have a statistically high chance that a time series with a clean peak will be collected, the time series collected must be relatively long. The 100 millisecond update interval is also dependent upon the processing time required for the sound localizing algorithm, and is limited by the time required for the motorized platform to be able to respond to a change in instruction. The time between updates must be a balance of these three factors. For instance, if the sound localizing algorithm can process a 100 millisecond data array in parallel with the acquisition of the next 100 millisecond data array, and if the motorized platform can be designed such that from a stationary position it can reach travelling speed in 100 milliseconds, then this is a suitable update interval.

The *data format* of the output of the sound localizing algorithm must be compatible with the control algorithm for the DMPS. It is described here so that output of any new sound localizing algorithm implemented in the future can be easily formatted to be compatible with the

existing control algorithm.

3.2.4 Stage 4 -- Directional Microphone Positioning System Design Specifications

The fourth stage is the DMPS. There are nine specifications for this stage. They are: total range of system, time allowed for travel through total range, overshoot, settling time, travelling speed, tracking speed, error on feedback signal, mean positioning error, and maximum positioning error.

The *total range* of the system is the angular range of motion through which the directional microphone can rotate. For a personal system, a full 360° range is not necessary because the voice of the user need not be amplified. A 270° rotation gives a "dead zone" which can be occupied by the user. A 180° range was chosen as the lower limit for this specification, as this would allow coverage of the entire half space in front of the user. In addition, limiting the positioning system to less than one full rotation of the directional microphone simplifies the feedback and control system by allowing the collection of absolute position information instead of change of position. Finally, in a multiple rotation system, the wires from the directional microphone would become tangled if slip ring electrical contacts were not used to connect the directional microphone to the stationary base of the device. The use of slip ring electrical contacts can add noise to the signal transmitted from the directional microphone. For ASLAD it is important to minimize any noise in the signal delivered to the hearing aid.

The *maximum time allowed* for the directional microphone to move through its full range of travel was chosen as 1 second. By timing the reading of text aloud, an estimate of the rate of speech was obtained. An average of three to four words were spoken per second. This means that less than one sentence would be lost in the time taken for the directional microphone to move to a new target. While it would be preferable not to lose any speech in the process of

moving the directional microphone, in most cases the loss of less than a sentence would not result in an inability to follow the conversation. However, in a fast paced volley of conversation this delay may prove to be a problem. This potential problem will be addressed in the discussion section of chapter 5.

The *maximum overshoot* of the directional microphone was chosen as $\pm 10°$, with a preferred overshoot of only $\pm 5°$. A negative overshoot represents an undershoot (shortfall). Directional microphones are rarely so directional that an overshoot of 10° would result in a loss of the desired signal during the overshoot. This allowance on the overshoot gives some flexibility in the control algorithm by allowing some oscillation about the desired position, while keeping the desired target within the range of the directional microphone.

The *settling time* relates to the rotational velocity of the directional microphone rather than the position. This parameter is closely tied to the time-between-updates parameter of stage three, and is set at the same value of 100 milliseconds. The DMPS must be able to respond to one solution of the sound localizing algorithm before the next is issued. This value is measured as the time elapsed between the application of a driving voltage to the motor for the system at rest to the time when a steady rotational velocity of the directional microphone is attained. This value is used because the response time for a system at rest will in general be the larger than that for a system in motion. This gives a worst case measurement for the settling time of the DMPS. While it is desirable to make the positioning system as efficient as possible, this specification could be relaxed in the event that the time-between-updates specification is increased.

The *travelling speed* is the total-range-of-system divided by the time-allowed-to-travel-through-total-range parameters. It is an estimate of the average rotational velocity of the directional microphone.

The ability to track a moving target was not chosen as a design requirement for ASLAD.

To track a moving target requires a much more complicated control scheme, likely involving some sort of predictive control method to anticipate the motion of the target. In the case of multiple sources, it also requires the ability to separately identify or characterize each target if the device is not to be easily fooled. A control algorithm capable of tracking moving targets was considered to be too costly in terms of processor time to be implemented in ASLAD. However, even with a point-to-point positioning system, reasonable tracking may be possible for slow moving targets. This specification is an estimate of the *maximum tracking speed* of the directional microphone based on the performance of the completed device.

The *error on the feedback signal* is the error allowed on the final position of the directional microphone. The allowable positioning error chosen is half of the allowable overshoot. The $\pm 5°$ allowable range about the desired position will help to minimize the motion of the directional microphone by minimizing the number of oscillations about the desired position required to settle within the acceptable range. While this is a relatively large error allowance, an error of five degrees in positioning would not have any negative effect on the amplification of a correctly localized target source. The two categories that follow this entry in Table 3.1, mean positioning error and maximum positioning error, are listed in anticipation of testing to be done on the DMPS once constructed.

3.3 Mechanical Considerations

There are a number of issues regarding the mechanical design that had to be considered in the construction of ASLAD. Some of these considerations, such as robustness, are general, and common to most mechanical designs. Others are specific to the application of assistive listening devices, such as silent operation and portability. In this section, three critical factors influencing the design are examined. They are the sound localizing algorithm resolution, size

and robustness of the device, and the need for a silent positioning system for the directional microphone. These three factors are the central to the physical appearance and mechanical design of ASLAD. A detailed summary of the construction of ASLAD follows.

3.3.1 Sound Localizing Algorithm Resolution and Mechanical Design

As described in section 2.4, the resolution of the sound localizing algorithm is dependent on the position of the sound source (*target*) relative to the pair of sensing microphones. Referring back to Figure 4(a), for a pair of stationary omnidirectional microphones, the localization resolution is the best for targets at position **A**, and worst for targets at position **C**. Figure 5 shows the change in resolution with target position for four different sampling rates. Even for high sampling rates, the resolution at and about position **C** is quite poor. This suggests that to maintain optimal resolution over the entire range of operation of ASLAD, the omnidirectional microphones should be rotated with the directional microphone to point at the identified target. Depending upon the angular separation of the target and the position of the directional microphone, the initial localization may be a rough estimate, but as the directional microphone and omnidirectional microphones were rotated towards the target, the resolution of subsequent localizations would improve until the directional microphone was successfully aimed at the target. This possible configuration is represented in the conceptual design shown in Figure 1 by dashed lines. A dashed line connecting the directional microphone to the omnidirectional microphones indicates that they would move together. The position of the directional microphone / omnidirectional microphone combination would also be required for the sound localizing algorithm in order to calculate target position. This is shown using a dashed line connecting the position sensor to the sound localizing algorithm in Figure 1. While the highest accuracy for a given sampling rate could be obtained with this configuration, there are two serious problems

associated with it.

The first problem with the rotating omnidirectional microphones is that the results of the sound localizing algorithm are not unique for a two microphone system. In section 2.4 it was noted that there are two possible target positions that could satisfy each solution of the sound localizing algorithm. These two target positions are symmetrical about the axis connecting the omnidirectional microphones, as shown in Figure 4(c). For the rotating omnidirectional microphone configuration, either the assumption that all targets lie in a range of less than 180° must be made, or a third omnidirectional microphone must be used in the determination of direction.

Limiting the range of operation of the DMPS would result in a reduction of the number of situations in which ASLAD would be useful. Targets outside of the range of operation would cause the directional microphone to be incorrectly positioned within the operating range due to the problem of the non-unique solutions obtained by the omnidirectional microphone pair.

For two microphones mounted on their side (pointing in the same direction as the directional microphone), there can be no distinction made between a low intensity sound in front of the microphones and a high intensity sound behind. An extra microphone pointed in the opposite direction to the pair of sensing microphones could be used to monitor sound intensity and determine whether a sound source was in front of or behind the pair of sensing microphones. There would have to be some shielding between the forward and backward facing microphones in order to ensure a measurable intensity difference between the microphones. The comparison of intensity levels using a third omnidirectional microphone could be used to derive a unique solution to the sound localizing algorithm, enabling operation over a full 360° range.

While the first problem with the rotating omnidirectional microphone configuration can be overcome with the addition of a third microphone and some modification to the sound

localizing algorithm, a second problem exists. Namely, it would require moving a rather unwieldy structure of arms with microphones mounted on the ends. The larger the separation between the two microphones, the higher the resolution of the sound localizing algorithm. Because a personal assistive listening device must be portable, there is a limit to the separation that can be used. For ASLAD, a separation of 20 cm was selected. A rotating structure with a 20 cm diameter would be flimsy, prone to damage, and visually distracting. It would require substantial dismantling for transport. Such a structure could be designed to be robust, but it would almost certainly have an increased weight and inertia. The increased inertia would result in greater difficultly controlling the positioning system to operate within the defined specifications. It would likely require a larger motor, and would increase power requirements.

With the rotating omnidirectional microphone configuration, the need for a third omnidirectional microphone has been established. Having made the transition to a three omnidirectional microphone system, the opportunity to explore other configurations of three (or more) omnidirectional microphones exists. With a larger omnidirectional microphone array, the need to rotate the omnidirectional microphones may be avoided. For instance, a device with a cylindrical base could have six, eight, ten, or more omnidirectional microphones mounted around the side of the cylinder, forming a ring. The base would shield opposing microphones from each other, creating intensity differences between the different microphones for a given signal. The intensity of signals collected from the microphones could be compared to determine general direction, and then the sound localizing algorithm could be applied to a pair, or pairs, of microphones on the appropriate side of the device. While this may work, it requires a significant increase in hardware requirements for the additional microphones and associated circuitry, filters, and A/D converters. To keep the hardware requirements and processing time to a minimum, the number of omnidirectional microphones must be kept to a minimum.

Consider the case of a <u>three omnidirectional microphone system</u>. By mounting the microphones at the vertices of an equilateral triangle, facing upwards, there are three possible combinations of two microphones. Figure 8 is an overhead view of a three omnidirectional microphone configuration. The three omnidirectional microphones are numbered 1, 2, and 3 in Figure 8, and the combinations can be referred to as the 1-2, 2-3, and 3-1 combinations. Furthermore, each of the three pairings has two regions in which the localizing resolution is high. Referring back to Figure 4(c), these are the ranges about positions **A** and **A'**. So for three microphones, six ranges of high localizing resolution are available. If the microphones are arranged in an equilateral triangle, these ranges will be equally spaced. To have coverage of the entire 360° field with six ranges of equal size, each range would have to span 60°.

In Figure 8, the angles $\phi_{1\text{-}2}$, $\phi_{2\text{-}3}$, $\phi_{3\text{-}1}$ are shown, defined the same way as in Figure 4(a), and a new angle ψ is defined. The angle ψ is an azimuthal angle for the 3 omnidirectional microphone configuration, which will be used henceforth in this script. In a three omnidirectional microphone system, results must be converted to some single scale to be able to keep track of them and to be able to produce a single value for transfer to the DMPS. The angle ψ is defined with $\psi = 0°$ behind the device because this coincides with the dead zone where the directional microphone cannot travel. By defining ψ in this way, it is continuous and positive throughout the operating range of the device.

The six equal 60° ranges can be given in terms of the angle ψ as: $30° < \psi < 90°$, $90° < \psi < 150°$, $150° < \psi < 210°$, $210° < \psi < 270°$, $270° < \psi < 330°$, and [$330° < \psi < 359°$, $0° < \psi < 30°$]. The highest resolution for each omnidirectional microphone combination is obtained for $\phi_{x\text{-}y} = 0$ (where $\phi_{x\text{-}y}$ represents the angle ϕ for any one of the omnidirectional microphone pairs), so for the 60° range given by ($-30° < \phi_{x\text{-}y} < 30°$) the lowest resolution is at $\phi_{x\text{-}y} = |30°|$. From Figure 5, this resolution is $\pm 8°$, $\pm 5.8°$, $\pm 5.4°$, and $\pm 2.5°$ for sampling rates of 15, 20, 22,

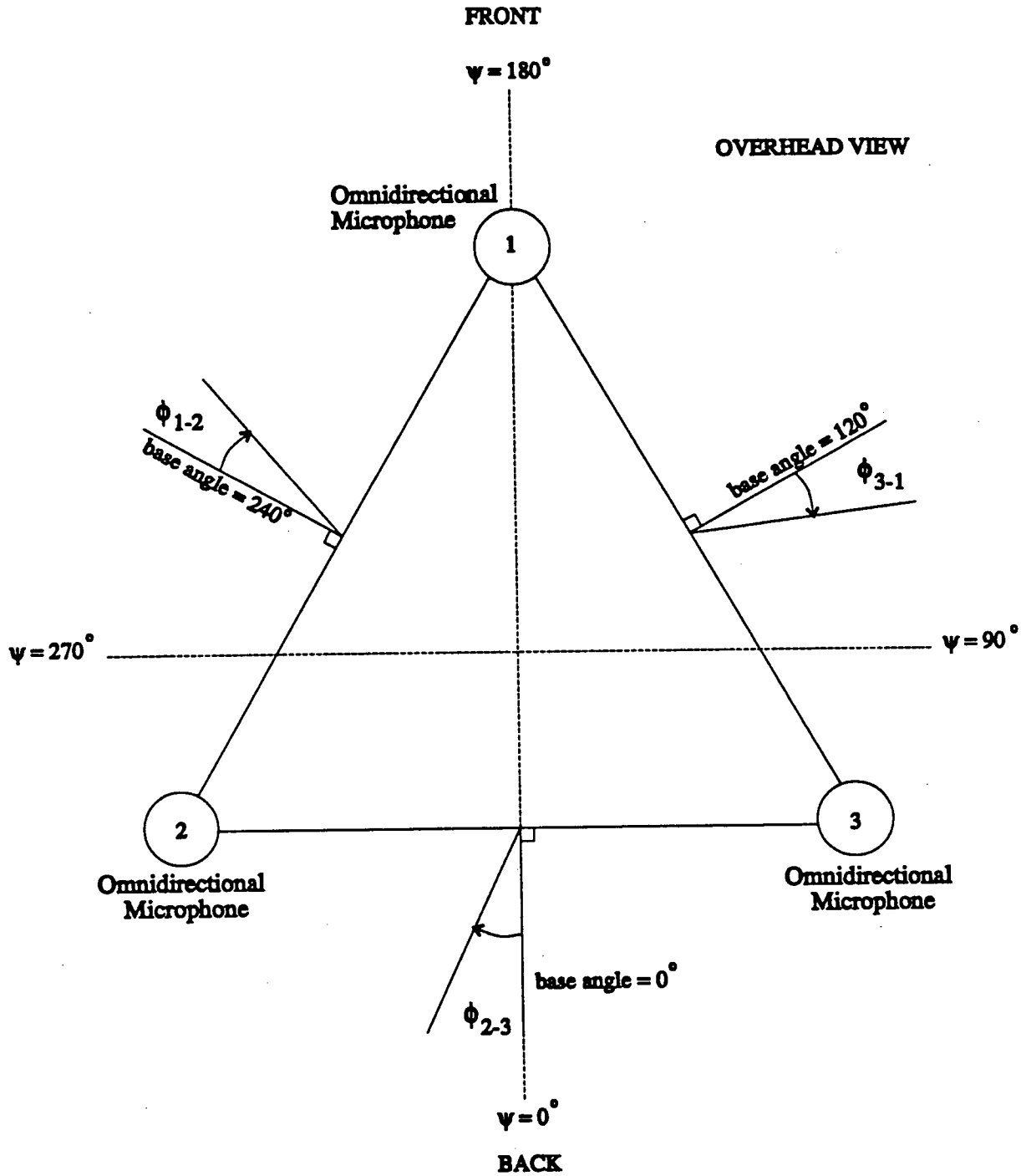


Figure 8: Overhead view of a 3 omnidirectional microphone configuration. A separate target angle ($\phi_{x\text{-}y}$) is defined for each omnidirectional microphone pair, and is shown only for the "outward" facing side of the triangular array to simplify the drawing. The *base angle* is the $\phi=0°$ angle for each omnidirectional microphone pair, converted to the ψ scale. A solution to the sound localizing algorithm is determined for each omnidirectional microphone pair. The results from each of the 3 pairs are compared to determine a unique solution. The final target position is converted to a solution in terms of the angle ψ, as defined on the drawing.

and 44 kHz respectively. So for a sampling rate of 20 kHz or above, target sources could be localized with a worst case resolution of less than $\pm 6°$ over a full 360° field. With this configuration, the sound localizing algorithm could be applied to each of the three combinations of two omnidirectional microphones without modification. By comparing the results of the sound localizing algorithm from each of the three pairs, a target position can be determined without ambiguity.

Before selecting a three omnidirectional microphone system such as the one described in the previous paragraph, the benefits of a <u>four omnidirectional microphone system</u> were also examined. Six possible combinations of two omnidirectional microphones exist for a four omnidirectional microphone system, but as shown in Figure 9 two of the combinations do not provide unique data. In Figure 9, solid lines connecting microphone pairs indicate combinations providing useful data. Pairs connected by dashed lines are parallel to other microphone pairs, and would thus give duplicate information. Again, with two regions of high resolution for each pair, the 360° field can be divided into eight equal ranges of 45° each. For the 45° range defined by ($-22.5° < \phi_{x-y} < 22.5°$) the lowest resolution is at $\phi_{x-y} = |22.5°|$. From Figure 5, this resolution is $\pm 7.4°$, $\pm 5.4°$, $\pm 5.1°$, and $\pm 2.35°$ for sampling rates of 15, 20, 22, and 44 kHz respectively. The improvement in the worst case resolution over the three microphone configuration for the identified sampling rates is 0.6°, 0.4°, 0.3°, and 0.15° respectively. For a sampling rate of 20 kHz or above, the improvement in worst case resolution is less than half a degree. The cost for this improvement is a further increase in hardware, and an increased processing time.

Once the sound localizing algorithm had been programmed, the <u>processing time</u> required was measured for both two and three microphone versions. The two microphone version required 44 milliseconds, and the three microphone version required 77 milliseconds. Based on the time

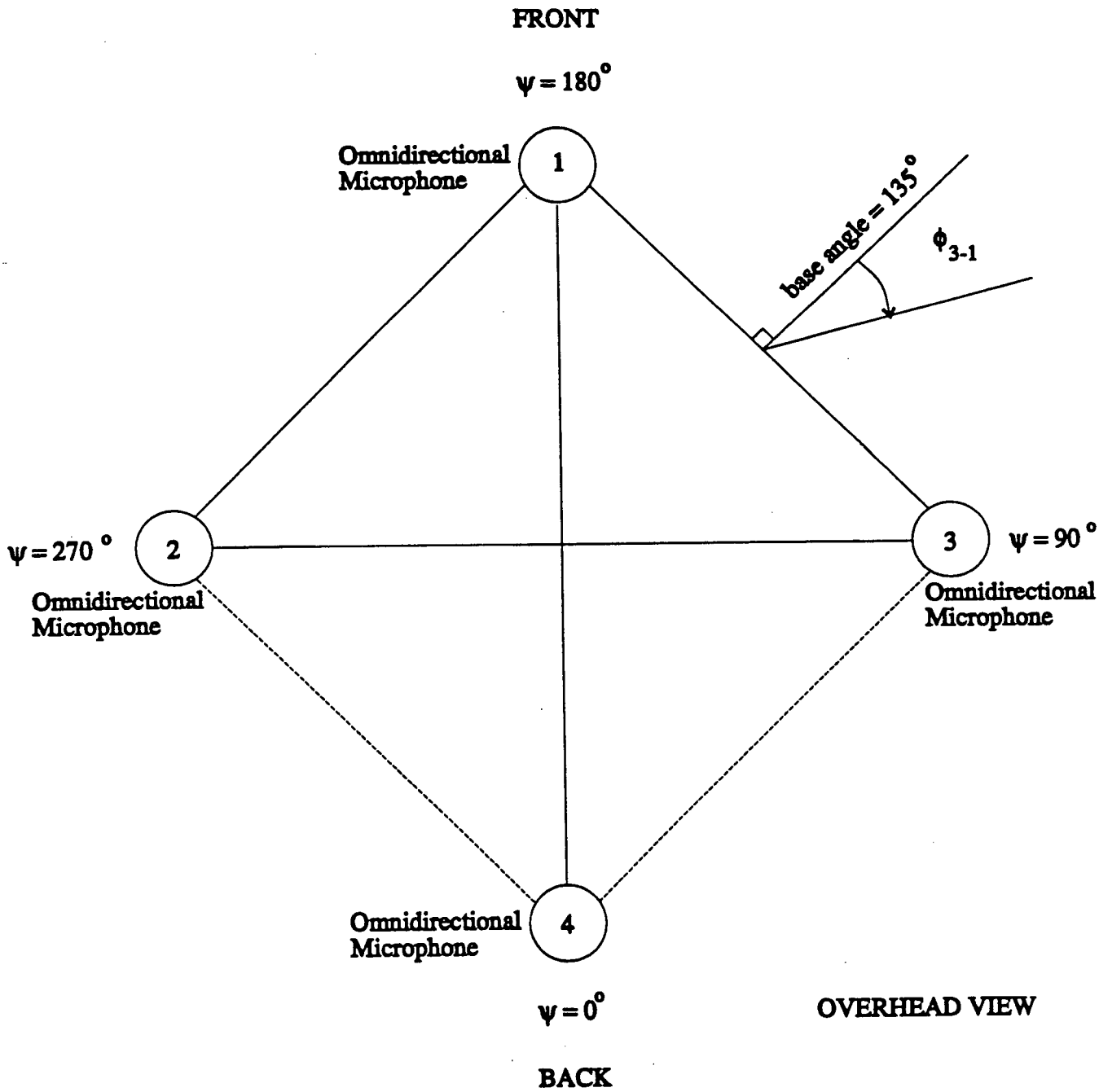


Figure 9: Overhead view of a 4 omnidirectional microphone configuration. Same operating principle as the 3 omnidirectional microphone design. Only one example of the omnidirectional microphone pair target angle ($\phi_{3\text{-}1}$) is shown. Solid lines connecting 2 omnidirectional microphones indicate a pair providing useful information. Dashed lines connecting 2 omnidirectional microphones indicate a pair providing duplicate information (eg. pairs 1-2 and 3-4 will yield the same information about the target position).

required for each of the different steps in the sound localizing algorithm, a fourth microphone would result in an additional processing time of approximately 26 milliseconds, for a total processing time of 103 milliseconds. We decided that additional costs associated with both hardware and processing time requirements were not justified for an incremental improvement in worst case resolution of less than half a degree.

The arrangement of three omnidirectional microphones in a triangle increases the size of the base of ASLAD, but vastly reduces the size and weight of the rotating portion. In addition, this design is more robust, and less prone to damage. By minimizing the presence of the moving components, the device will also be less distracting to those using it. The triangular omnidirectional microphone array was chosen as the final design for the input stage of ASLAD. Figure 10 shows the variation in localizing resolution over a full 360° range for different sampling rates.

A report of one other sound localizing system using a triangular input array was found in the literature. In a paper by Sugie *et al.*(1988), an array of three microphones at the vertices of an equilateral triangle of side 13.5 cm, for use in a sound localization experiment, was described. However, for their experiment, the microphone array was manually rotated while one or two stationary target sources were presented. The data collected was processed off-line using an algorithm based on the IPD.

3.3.2 Silent Operation and Mechanical Design

There are three reasons why it is critical that the DMPS operates silently. The first is that any noise made by the positioning system while in motion would be picked up by the sensing microphones. This additional noise source could negatively affect the performance of the sound localizing algorithm. Since any noise made by the positioning system would originate from a

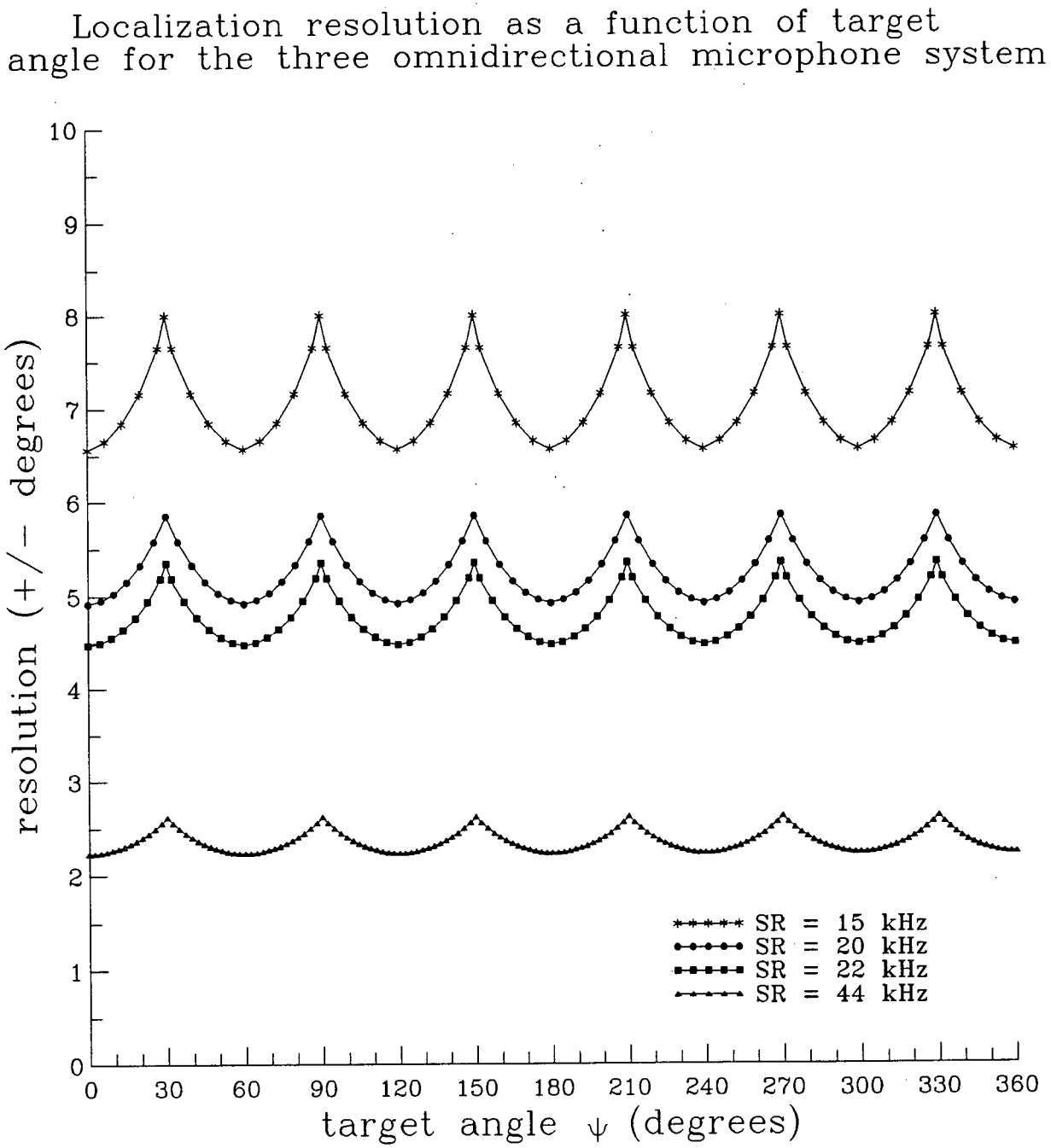


Figure 10: Localization resolution as a function of target angle (ψ) for the 3 omnidirectional microphone configuration. Localization resolution for four different sampling rates (SR) are shown.

fixed position relative to the omnidirectional microphones, noise from it would always result in the same localization. This could be compensated for in the software, however, there are two other reasons why silent operation would be necessary for a practical device.

The second reason why the DMPS must operate silently is that any sounds made by it may be picked up by the directional microphone. Any noise would be attenuated both by the casing that holds the positioning system, and due to the location of the positioning system relative to the directional microphone. However, any noise added to the desired signal makes interpretation more difficult for the hard of hearing individual, and should be avoided if at all possible.

Finally, the third reason is that assistive listening devices are generally used in the presence of people with normal hearing. Already in this design we have the visual distraction of the rotating directional microphone. While it may be quite easy for people to learn to ignore this, it would be very difficult if it also clicked, whirred, and buzzed as it moved around.

DC motors can operate silently, but they generally must be geared down to a suitable operating speed. As shown in equation 2.13, this reduces the torque required to turn the load shaft, and allows the use of a smaller motor. The gears are often a source of noise. Direct drive motors exist, but are most commonly used for large scale industrial applications. Two options for a silent positioning system were considered: linear actuators with spring returns, and belt-driven systems.

One possible silent positioning system was an arrangement which balanced a linear actuator against a spring return system. The plunger of the linear actuator would be attached to a rotating platform (on which the directional microphone would sit) such that as the plunger moved in, it turned the platform in one direction, and as the controlling voltage on the actuator was reduced, the spring would pull the platform back in the other direction. In such a system,

the controlling voltage for the linear actuator would be directly related to the position of the platform, thus providing absolute position feedback information. This system would be quite simple to control. It was decided against on the basis that, for all but one resting position, power would be required to maintain a fixed position. If one actuator-spring combination was used, this resting position would be at one at one extreme of the range of travel. If two actuator-spring combinations were used in opposition, then the resting position would be at the midpoint of the range of travel.

The other candidate for a silent positioning system that was considered, and eventually chosen, was a belt-driven platform. For rotation about a vertical axis, a belt-driven platform can be stationary at any position without power consumption. The belt connects a small pulley on the motor shaft to a larger pulley on the platform shaft. It performs the same function as a set of gears, but is quiet. One drawback of the belt-driven system is that it is prone to slippage. A toothed belt and pulley combination would have been a good solution to this problem, but was not available in the small size required here. Instead, the position feedback has been designed to give absolute values, so that if slippage occurs, the angular position of the platform will still be known.

3.3.3 Construction Details

Figures 11(a) and 11(b) show dimensioned drawings of the exterior of ASLAD. Figure 11(c) is a photograph of ASLAD. The base of ASLAD is a steel box with dimensions of approximately 26 cm x 26 cm x 6 cm. Along the back side, plugs have been installed for connections to a DSP board, a power supply, output for the directional microphone and an on/off switch. This box houses the motor and platform assembly, the driving circuitry for the motor, the amplifier circuits for the omnidirectional microphones, and the circuitry for the directional

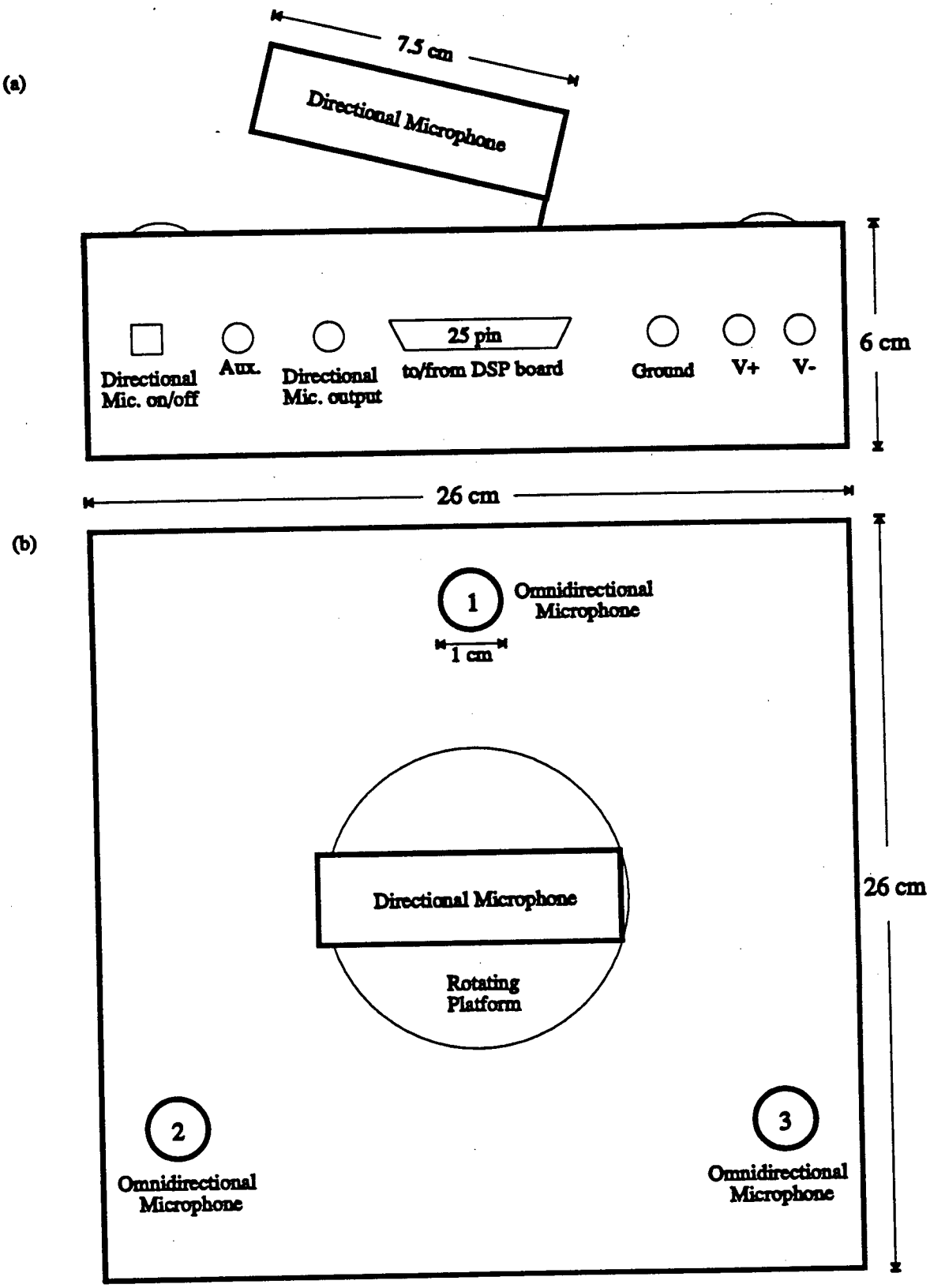
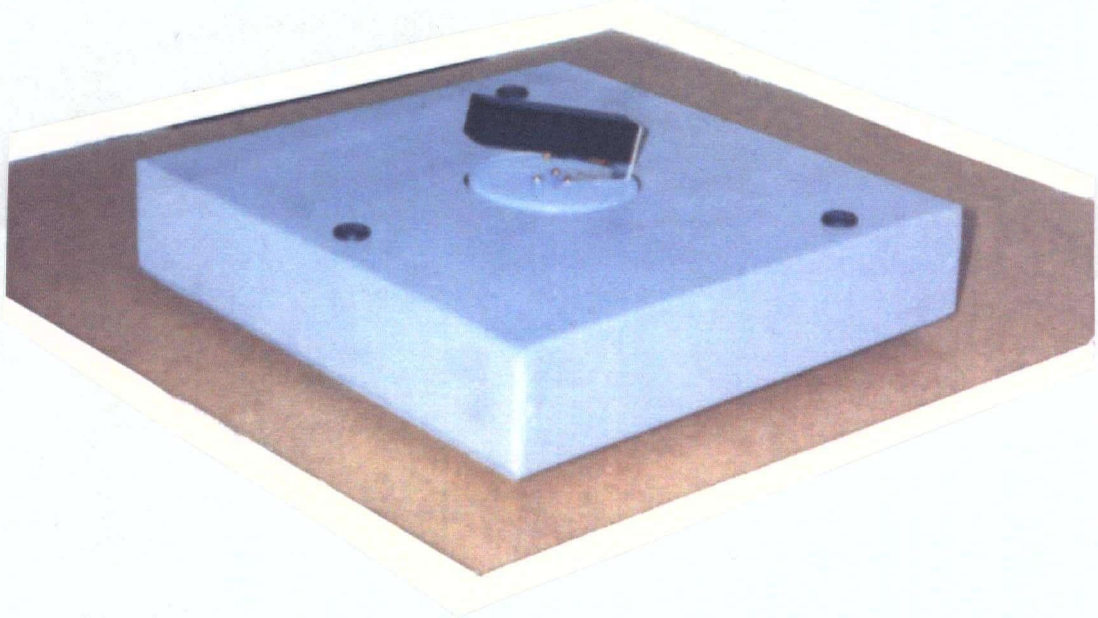


Figure 11: Dimensioned drawings of ASLAD: (a) side view and (b) top view; (c) photograph.



microphone. The top surface has three 0.9 cm diameter holes centred at the vertices of an equilateral triangle of side 20 cm, and a 7.8 cm diameter hole centred at the geometric centre of the equilateral triangle. The equilateral triangle is oriented such that one side is parallel with the back edge of the housing. The three 0.9 cm holes hold the omnidirectional microphones. These are held in place with a friction fit using 3/8" rubber grommets. The 7.8 cm diameter hole is for the platform that holds the directional microphone.

The omnidirectional microphones are Archer™ brand (distributed by Radio Shack) electret condenser mike elements. They have a nominal output impedance of 1 kΩ, nominal supply voltage of 4.5 volts, and sensitivity of -65 dB ±4 dB (0 dB reference 1 V/μbar at 1 kHz). The frequency response is basically flat in the range 30 Hz to 3 kHz, with a slightly reduced sensitivity below 30 Hz, and a slight peak in the frequency response at about 5 kHz followed by a roll off of approximately 10 dB/decade. Any small omnidirectional microphones would have sufficed for this application. These were readily available, inexpensive, and easily mounted into the ASLAD case with no modification to their packaging required.

As explained in section 3.3.2, a belt drive was chosen for use in the positioning system. Had the goal of the project been to develop a design for mass production, the positioning system would have had to be built from scratch using commercially available and/or custom components. This was not the goal of the project -- the device to be built was a prototype. Because the design and construction process can be lengthy for a mechanical device, and because only one device was to be built, an effort was made to locate a partially constructed positioning system that could be modified to suit the requirements of ASLAD. A suitable belt-driven platform and DC motor combination was found to be the drive from an audio cassette player.

The motor, drive belt assembly and platform axis were taken from a micro-cassette tape recorder. Once unnecessary parts were removed, this provided a ready made framework with a

small DC motor connected to a metal pulley via a rubber belt. The platform onto which the directional microphone would be mounted could be attached to the metal pulley (Figure 12). Using this pre-fabricated framework ensured that the drive belt was properly tensioned.

The motor is a generic permanent magnet DC motor, which operates equally well in either the clockwise or counterclockwise direction. The metal pulley is approximately 3 cm in diameter, and is solid cast metal. This forms the base for the platform on which the directional microphone sits (Figure 12). The drive belt is connected to the motor shaft by another small pulley, which is approximately 3 mm in diameter. The ratio of the diameters of the motor pulley to the load pulley is therefore 1:10.

In order to provide position information, the 3 cm pulley was removed from the framework, and the resistive element and contact brush assembly from a carbon film potentiometer were mounted between the pulley and the framework. One side was attached to the pulley, and one fixed to the framework.

Four holes were drilled in the pulley. One was positioned at the edge of the pulley such that it is aligned with the front of the directional microphone. This was fitted with a small pin, which acts as a mechanical stop at each end of the range of travel of the platform. The other three holes were arranged in a triangle, and were tapped. These hold the screws that support a 7.5 cm diameter rigid plastic disk on which the directional microphone is mounted (see Figure 12).

A fixed array of microphones such as those described in the introductory chapter of this thesis could have been used for the directional microphone. However, building the array and the associated circuitry would have been very time consuming, and there was no obvious benefit to building a custom directional microphone if a suitable commercial product existed. As it turned out, an appropriate directional microphone did exist, and is in itself an assistive listening device.

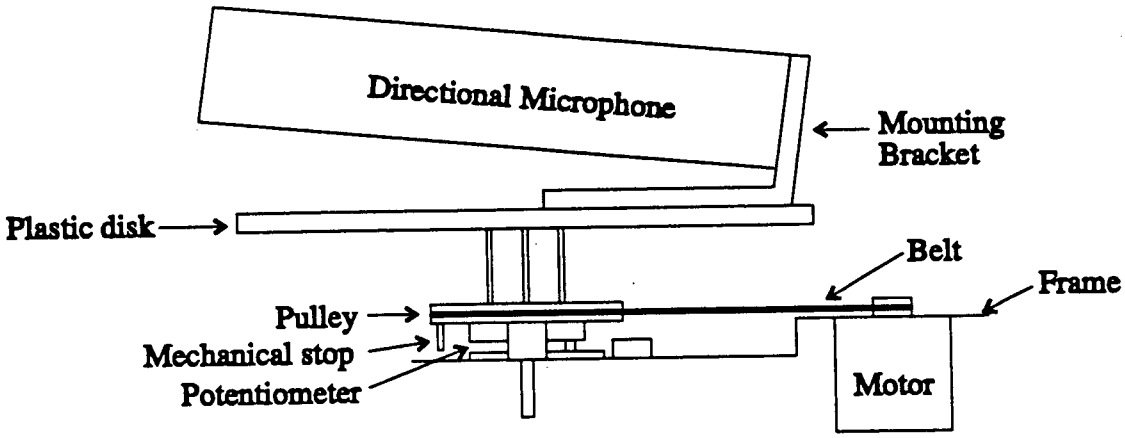


Figure 12: Side view of the internal framework of the directional microphone positioning system, which holds the motor and rotating platform assembly.

The directional microphone is a Phonak™ SDM (monaural) conference microphone. It is designed to plug directly into a particular brand of hearing aid, and is manually positioned by the user to point in the desired direction. This microphone has an acceptance angle of approximately 60°. This value is measured at the -3 dB points on a polar plot of the microphone response (Figure 13). The gain directly in front of the microphone (along its longitudinal axis) is approximately 15 dB greater than that to the side of the microphone (along the transverse axis), and is approximately 25 dB greater than behind the microphone. It has an effective range of approximately 4 meters. It is very compact, measuring 7.4 cm in length, but has directional characteristics comparable to microphones measuring 30 cm or more in length. The body of the microphone is divided into two equal parts, one housing the microphone assembly, and the other housing the electronics and battery. The microphone weighs 74 g total, but most of this weight is associated with the electronics, which has been removed from the platform. The portion of the microphone assembly remaining on the platform weighs 16 g. It would be difficult to find a directional microphone much smaller and lighter than this. The inertia of the extra rotating platform and microphone is significant compared to the existing pulley, but when translated back to the motor, it is still less than that of the motor itself. The combined effects of additional friction and inertia are probably no more than that of the reel of audio tape that it was initially designed to turn.

The directional microphone is mounted, using an aluminum bracket, at an angle of 11°, approximately the same as it would have been had the directional microphone been being used in its originally intended fashion. This angle is such that if the device is set on a table, and the users are seated at the table, the microphone will be pointed at face level at a distance of 2 meters. Users seated closer to or further from the directional microphone would still fall within the acceptance angle of the directional microphone. There is a slot cut in the plastic disk to

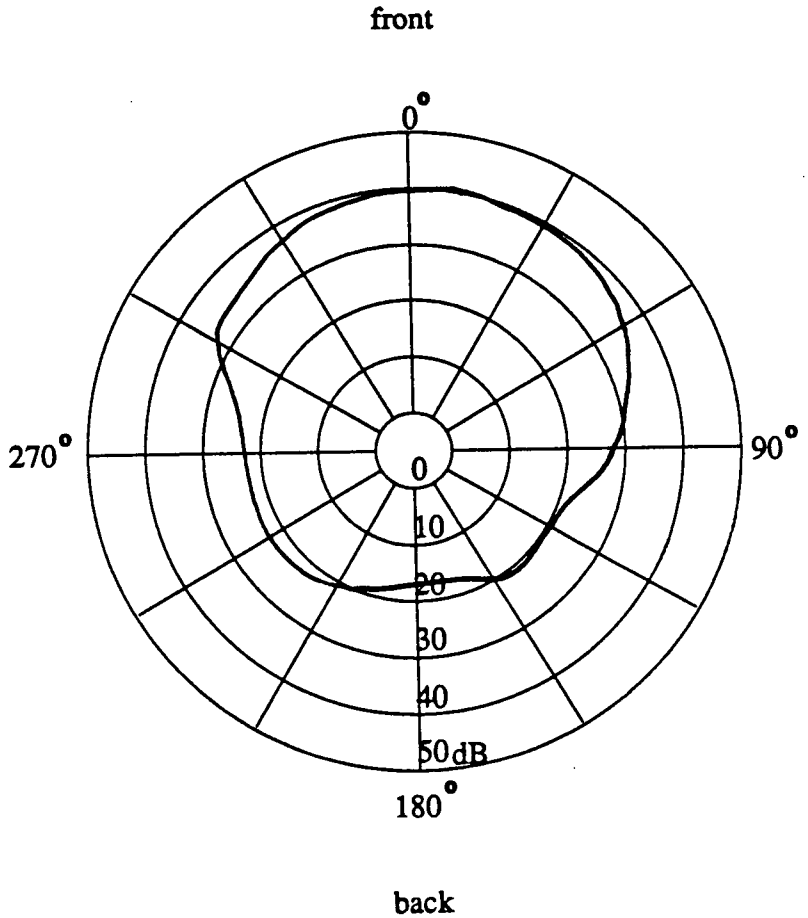


Figure 13: Gain characteristics of the Phonak SDM directional microphone for 1000 Hz test signal.

allow wires from the directional microphone to be fed down into the box. The whole platform assembly is mounted such that the plastic disk is flush with the surface of the box. It was necessary to raise the platform for two reasons. One was to allow for adequate clearance above the motor shaft and other protrusions on the framework that the motor and pulley were mounted to. The other was to allow for the wires from the directional microphone to be fed down into the box.

A preferable way to feed the wires from the directional microphone down into the box would have been through a hollow axle, which would have acted as a conduit for the wires. This would minimize the bending of the wires as the microphone rotates, and thus reduce the effects of fatigue on the wires, as well as minimize any capacitive effects caused by bending the wires. The absence of a hollow axle was one of the limitations of modifying an existing belt-driven system for use in this prototype. However, it was decided that the benefits of using an available belt-driven system was such that the wiring of the directional microphone would be of secondary priority. As described in section 3.2.4, the use of a slip ring to bring the signal from the directional microphone down into the box was ruled out due to the noise that it may add to the signal.

3.4 Electronics

There were three different circuits built for ASLAD. They are shown in Figure 14. The control signal for the DMPS is digital-to-analog (D/A) converted and passed through a current amplifier circuit (Figure 14(a)). The D/A converter has a range from -2.5 to +2.5 volts, and the current amplifier is necessary to produce a driving signal with sufficient current to drive the motor. For this a complementary pair of medium power transistors was used in conjunction with an op-amp to obtain current amplification for both positive and negative driving signals. The

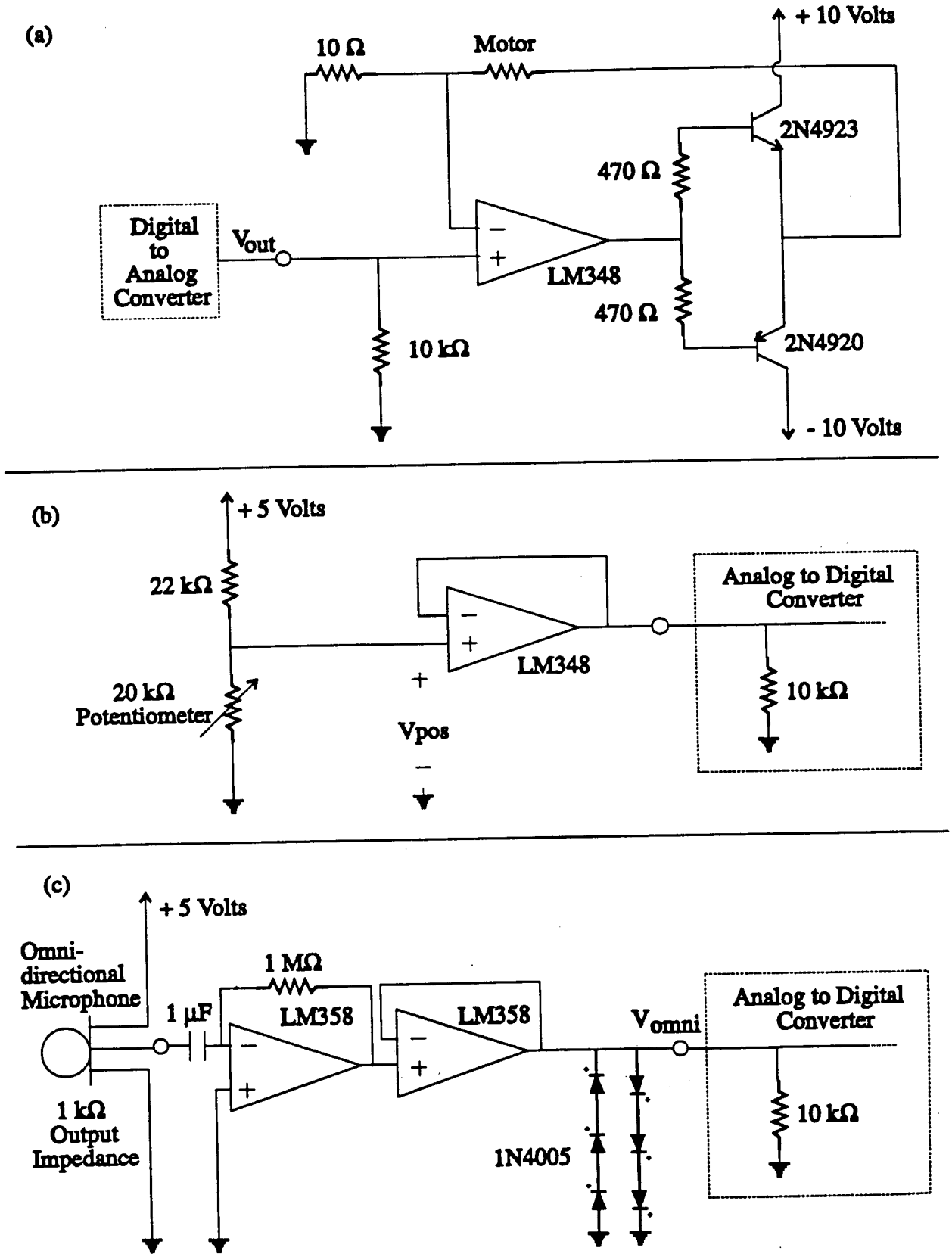


Figure 14: Circuits used in ASLAD: (a) current amplifier used to drive motor; (b) feedback circuit for position detection; and (c) amplifier and hard limiter circuit for the omnidirectional microphones.

circuit is based on a design in (Jung, 1974).

A simple voltage divider circuit is used to provide the position feedback. See Figure 14(b). A unity gain buffer is necessary as the input stages of the A/D channels have a 10kΩ resistor to ground. By using a buffer amplifier to reduce the output impedance of the voltage divider circuit to a very small value relative to the 10kΩ, an essentially linear scale is achieved. In this way, the accuracy achievable is constant throughout the range of travel of the directional microphone.

The maximum time delay between the arrival of a signal at one microphone and another is 583 microseconds (assuming a value for the speed of sound in air of 343 m/s). This maximum time delay would occur for a source originating at position **C** in Figure 4(a). For such small time delays, it is important to minimize the analog pre-processing of the signal. Analog circuits are subject to change in performance over time due factors such as drift in the value of the resistors used. Techniques such as digital filtering are immune to temperature differences, and invariant over time. In addition, each signal is subject to identical treatment. Change in the behaviour of one circuit relative to another could result in errors in the localization. If the transfer functions of two circuits changed relative to one another enough to shift the digitization of one signal by one sampling interval, the resulting error in localization would be about 5°. Small changes in the behaviour of the circuitry can quickly result in erroneous results, compromising the performance of ASLAD.

Figure 14(c) shows the amplifier circuit used for each of the three omnidirectional microphones. Basically, the raw signal is amplified, buffered for reasons described above, and limited at the output to approximately ±2.2 volts by a double string of diodes. This last stage is necessary in order to prevent damage to the A/D converters. The amplifier gain for the omnidirectional microphones is nominally 1000. This value was chosen experimentally. It gave

a signal magnitude close to the full range of 2.2 volts for normal conversation at a distance of approximately one meter. The sound localizing algorithm is designed such that differences in the gain of the three omnidirectional microphone amplifier circuits will not bias the results. The problem exists that as the distance between the target and the omnidirectional microphones increases, the intensity of the acquired signal decreases. The use of a variable gain amplifier to extend the range of useful operation was not suitable for use in this situation for two reasons. One is it would add to the complexity of the analog circuitry, which could be detrimental to the accuracy of the sound localizing algorithm. The other is that relatively long sound samples are used in the analysis of the target location (approximately 100 milliseconds). A variable gain amplifier generally has a much shorter reaction time, so the characteristic peaks required to obtain any useful information from the cross-correlation would be modified, if not lost.

An anti-aliasing filter was required for the omnidirectional microphone input stage. The first stage of each of the A/D converters is a third-order Butterworth active filter. The default -3 dB point for these low-pass filters is set to 34 kHz, but they are adjustable, and can be altered by changing a set of plug-in resistor packs. Because these filters preceded the sampling and digitizing stages of the converters, it was possible to use them as anti-aliasing filters. The value of the plug-in resistor packs required to set the roll off frequency of the filters to 10 kHz was not readily available, so instead the next lowest value was chosen. This gave a 7.25 kHz low pass filter without having to add an additional stage to the amplification circuitry. A lower bandwidth filter was chosen because the sampling rate was also lowered from the original 44.1 kHz during the course of the implementation of ASLAD.

The directional microphone used has its own signal conditioning circuitry, and is designed to connect directly to a particular brand of hearing aid. When connected to the hearing aid, the circuitry in the directional microphone automatically reduces the input from the hearing aid

microphones by 6 dB. For situations in which the background noise is high, there is a switch on the directional microphone that reduces the hearing aid input by 18 dB. The hearing aid performs the functions of hard limiting and signal processing, specific to the user's needs, on the signal from the directional microphone. For ASLAD, no treatment of the signal from the directional microphone was required.

3.5 Signal Processing

The program that controls the operation of ASLAD is implemented on a Spectrum signal processing platform. It has a Texas Instruments TMS320C30 DSP chip, which has 2 kilobytes of on-chip, zero wait-state, memory (RAM). It also has another 64 kilobytes of on-board memory (1 wait-state). Data is collected from, and sent out via, a 4-channel A/D, 2-channel D/A converter board. The program has three main components: data collection, the sound localizing algorithm, and the directional microphone control algorithm. The sound localizing algorithm and the directional microphone control algorithm are further broken down into subroutines. This provides easy modules to work with during program modification and debugging.

ASLAD is a new assistive listening device, and ideas from many different fields are brought together in this implementation. It was desirable that ASLAD be adaptable to future changes and modifications. It is for this reason that the processing is based on a DSP board. The program that runs the device has been written in a modular fashion, calling repeatedly on a set of functions to accept data, perform sound localizing and directional microphone positioning. This allows the sound localization algorithm used in this implementation to be removed and replaced with a new or different version without having to restructure the other components of the program. It is likely that the majority of changes made in future versions of this device would be in the sound localizing or control algorithms.

3.5.1 Data Input

Data is collected from the four A/D converters using an interrupt subroutine. The conversion is triggered by a counter set to give the appropriate sampling rate, and the interrupt subroutine is invoked at the end of each A/D conversion. Three channels are for the three input signals from the omnidirectional microphones, and the fourth channel is for the feedback signal for the directional microphone position. The interrupt subroutine can be enabled or disabled by writing different codes to a control register on the DSP card. The interrupt subroutine is enabled for the duration of the data collection period, then disabled during the execution of the subroutines comprising the sound localizing algorithm and the DMPS control algorithm. It is re-enabled at periodic intervals in between the execution of each of the many subroutines that comprise the sound localizing algorithm to acquired updated position information for execution of the DMPS control algorithm.

Once digitized, each data point of the signal from each omnidirectional microphone is rectified and compared to a threshold value before it is stored in an array. The threshold value is set at 0.36 volts, and any data point below that value is set to zero. This value was chosen by measuring the peak signal level in a reverberant environment with no target source present. It was a measure of the background noise in the reverberant environment. By setting a threshold, the rising edge of the waveform of a target source is much sharper. This provides an additional feature in the signal that may help improve the accuracy of the sound localizing algorithm. The threshold value is easily adjustable in software. If desired, the estimation of a threshold value could be incorporated into a start-up calibration routine in future implementations of ASLAD.

Once a 2000 point sound sample has been collected, the arrays are padded with zeros up to a length of 2048. Padding the ends of the arrays with zeros is a necessary step to prevent erroneous results from the convolution procedure used. It also brings the length of the arrays up

to a value of 2^n, in order that FFTs and inverse FFTs (IFFTs) can be performed.

The initial design value for the sampling rate was 44.1 kHz. This was lowered during the course of programming the DSP chip. The reason was that the DSP chip has only 2 kilobytes of high speed memory (Random Access Memory). To achieve the full benefit of the fast processing times of the DSP chip, the arrays being FFT'd and IFFT'd had to be loaded into the RAM. This meant that the longest array that could be used was 2048 samples. At 44.1 kHz, a 2000 point sound sample is only 45.3 milliseconds long. A 45 millisecond sound sample does not give a statistically high chance that a time series with a clean peak will be collected. By reducing the sampling rate to 20 or 22 kHz, a 2000 sample array of 100 or 90.9 milliseconds in length (respectively) can be collected, while keeping the processing time to a minimum. This is important as the signal processing is heavily dependent on the use of FFTs and IFFTs. If a 4096 point array had to be used to obtain a long enough sound sample, the processing time would likely be quadrupled. Of course, the cost of changing the sampling rate is a decrease in the localization resolution, but the worst case resolution due to the digitization of the target signal is still kept to within $\pm 5.4°$ for the 22 kHz sampling rate and $\pm 5.8°$ for the 20 kHz sampling rate. On the whole, the likelihood of acquiring a clean peak in the time series of the sound samples is doubled, the processing time is kept to a reasonable level, and the worst case resolution for the sound localizing algorithm is still under $\pm 6°$.

The first step in the sound localizing algorithm described in section 2.4 is to low-pass filter the data. For this algorithm, it would be possible to achieve the desired time domain resolution by using a lower sampling rate and subsequently increasing the time domain resolution via an interpolation method. This is sometimes referred to as *upsampling* (Oppenheim and Schafer, 1989). However, if a lower sampling rate were to be used, the bandwidth of the anti-alias filter would also have to be reduced. As one of the design requirements is to make ASLAD as adaptable as possible to the implementation of new algorithms, it is not desirable to limit the

input bandwidth.

3.5.2 Sound Localizing Algorithm

Figure 15 is a flowchart of the steps in the sound localizing algorithm. Each step is written as a separate subroutine. The control algorithm for the DMPS is executed in between calls to each of the subroutines in the sound localizing algorithm.

Each of the three arrays of data collected from the omnidirectional microphones have been rectified at the input stage, as described in the previous section. The next step is the low-pass filtering of the arrays to obtain the envelope of each signal. The arrays are digitally low-pass filtered using the FFT-based convolution method described in section 2.5.2, in which the sampled data is FFT'd, multiplied by the FFT of a filter function, then inverse FFT'd.

The filter function used in the sound localizing algorithm was provided by Zakarauskas[1]. The inset in the upper right corner of Figure 16 shows the normalized filter function which, in the time domain, would be convolved with the data collected from each omnidirectional microphone to achieve the effect of low-pass filtering. It was chosen, by Zakarauskas, from a selection of common windowing functions, such as the Hanning window and others based on the cosine function. The filter function was chosen based on experiments in which different filters were implemented in a two omnidirectional microphone version of the sound localizing algorithm. All of the functions tested shared the common characteristics that they were symmetrical about the zero bin, had a "Gaussian" shape, and resulted in the presence of very little ringing in the filtered data. The function shown was chosen because, of the filters tested, it gave the best results (i.e. the most precise localization estimates) for the two omnidirectional

[1] Filter function provided by P. Zakarauskas in personal communication.

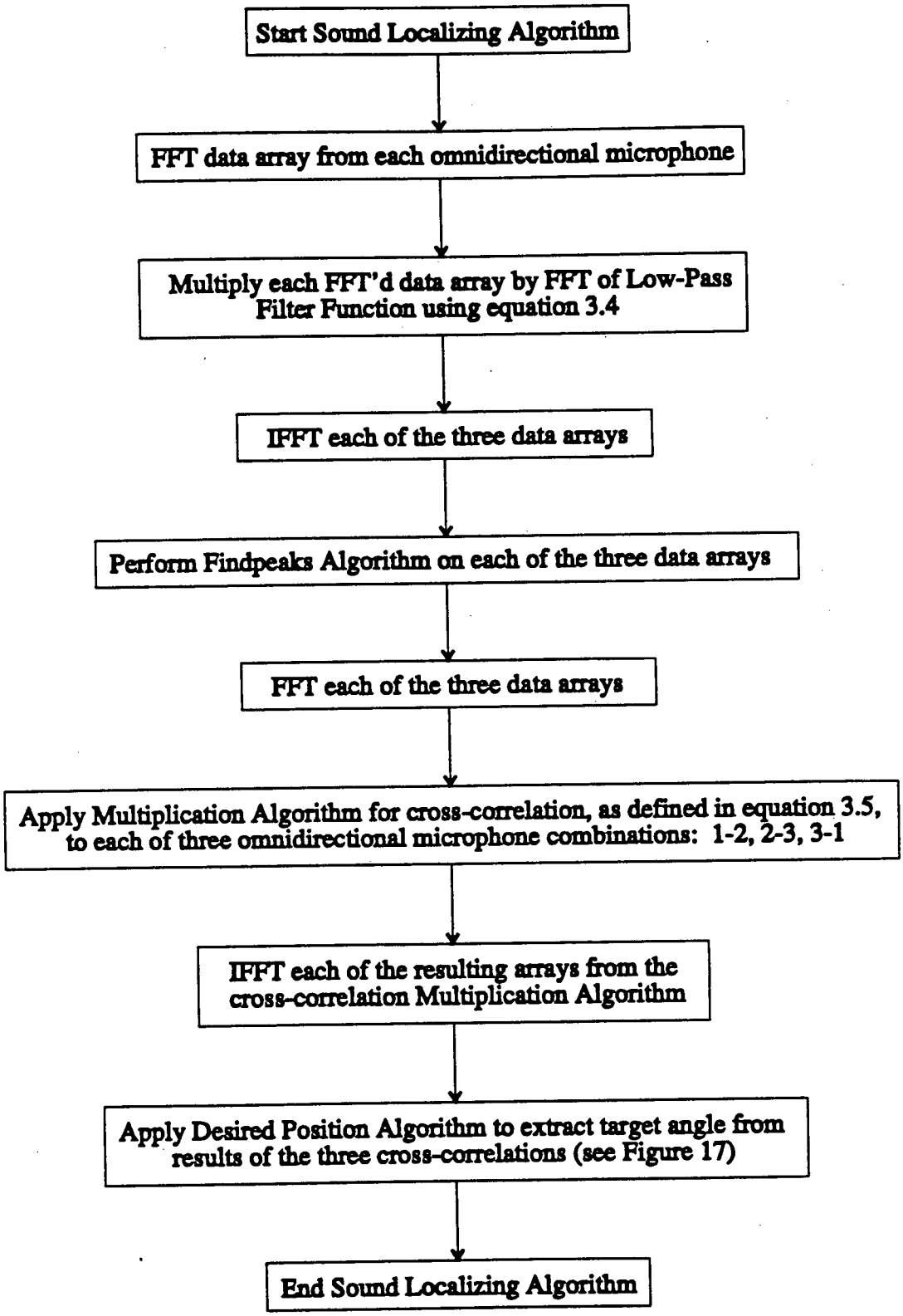


Figure 15: Flowchart of the sound localizing algorithm.

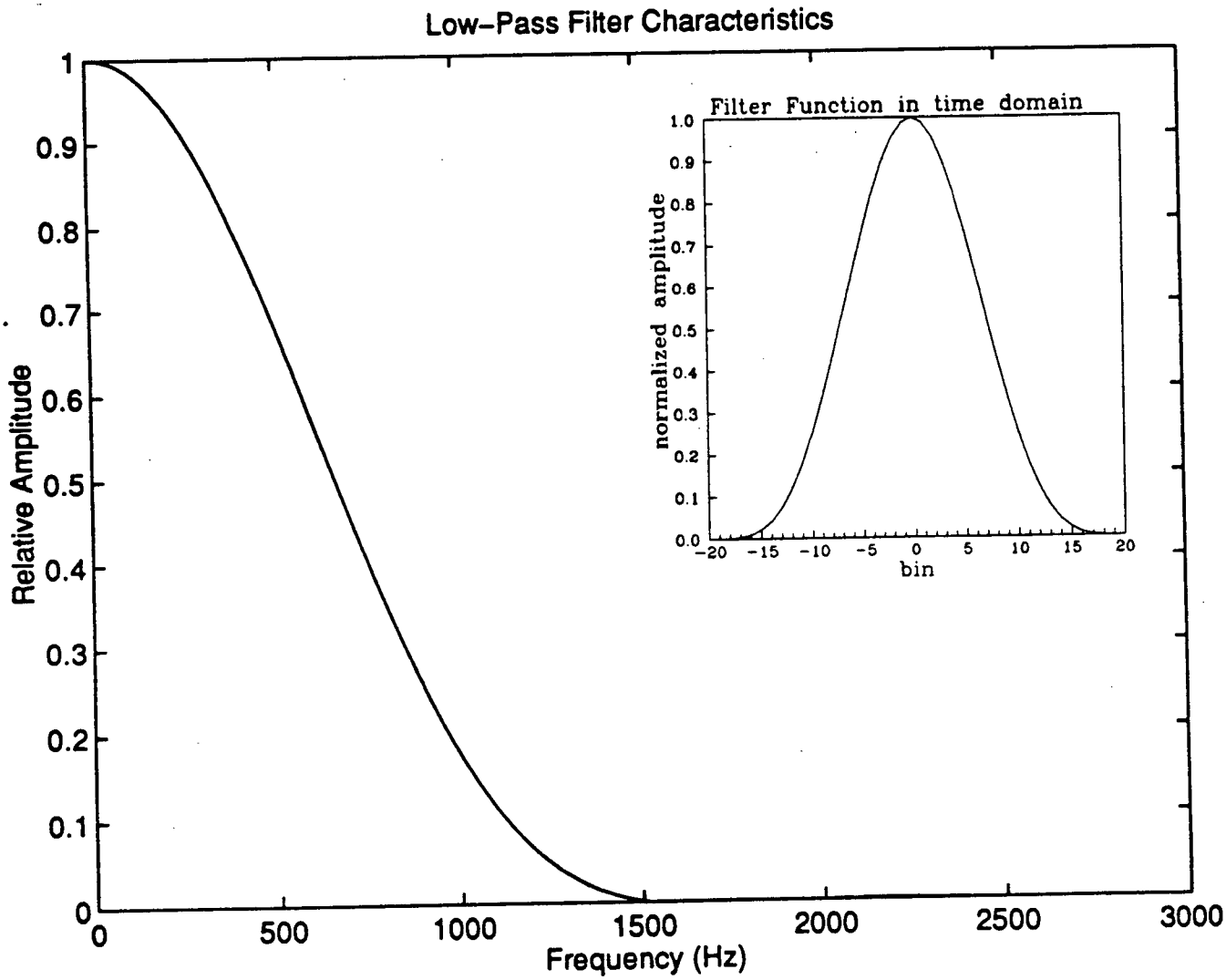


Figure 16: Digital low pass filter used to obtain the envelope of sound samples collected from the omnidirectional microphones. The main plot is the Fourier transform of time-based filter function shown in the inset.

microphone implementation of the sound localizing algorithm. As shown in the inset of Figure 16, the function is real-valued and is even (symmetrical about bin zero). The Fourier transform of a real-valued and even function is also real-valued and even. The main plot in Figure 16 shows the Fourier transformed convolution function for positive frequencies. The -3 dB breakpoint is at 425 Hz.

For the case of a 20 kHz sampling rate, the filter function was shown in the inset of Figure 16. The definition of the filter function is given in equation (3.1), presented in a 2048 point array, arranged in the order necessary to perform the convolution using the FFT method. This is a *wrap-around* order, where the negative half of the time based filter function in the inset of Figure 16 is stored at the end of the 2048 point array:

$$F(i) = \frac{\left[\cos\left(\frac{(i-1)\pi}{40}\right)\right]^4}{\sum_{N=1}^{2048} F(N)} \qquad (1 \le i \le 20)$$

$$F(i) = 0 \qquad (21 \le i \le 2028) \qquad (3.1)$$

$$F(i) = \frac{\left[\cos\left(\frac{(2048-(i-1))\pi}{40}\right)\right]^4}{\sum_{N=1}^{2048} F(N)} \qquad (2029 \le i \le 2048)$$

It should be noted that the length of this filter must be modified for other sampling rates if the same frequency response is to be obtained.

The algorithms used to perform the FFTs and IFFTs were downloaded from the Texas Instruments Inc. bulletin board. They are described in Papamichalis (1990), and are originally adapted from Sorensen *et al.* (1987). They are written specifically for the TMS320C30 DSP chip, and are optimized to perform the transforms in the minimum amount of time.

The Texas Instruments FFT algorithm packs the transformed data of a real array of length N and order [x(0), x(1) ,..., x(N-1)] into an array also of length N in the following order:

$$R(0),\ R(1)\ ,\ldots,\ R\left(\frac{N}{2}\right),\ I\left(\frac{N}{2}-1\right),\ldots,\ I(1) \tag{3.2}$$

where R(k) and I(k) represent the real and imaginary components of the complex number X(k). The data points of the Fourier transformed array which are not stored can be recovered from (3.2) by using the following relations:

$$\begin{aligned}
&X(0) = R(0) \\
&X(k) = R(k) + jI(k) \qquad\qquad k = 1,\ldots,\left(\frac{N}{2}-1\right) \\
&X\left(\frac{N}{2}\right) = R\left(\frac{N}{2}\right) \\
&X(k) = R(N-k) - jI(N-k) \qquad k = \left(\frac{N}{2}+1\right),\ldots,\ (N-1)
\end{aligned} \tag{3.3}$$

To perform the real IFFT (complex array in frequency which is real in the time domain) the data points are packed into the array in the same order as shown in (3.2). The multiplications for the convolution can be done without expanding the arrays. For a 2048 point array, the formulae are as follows (where Filter(i) is the Fourier transform of the filter function, data(i) is a Fourier transformed sound sample, and $data_f$(i) is the array in which the results are stored):

$$\begin{aligned}
&for\ i = 0\ to\ 1024\ : \\
&\quad data_f(i) = data(i) * Filter(i) \\
&for\ i = 1025\ to\ 2047\ : \\
&\quad data_f(i) = data(i) * Filter(2048 - i)
\end{aligned} \tag{3.4}$$

The multiplication for the cross-correlation is similar, and is given by:

$$
\begin{aligned}
&data_{corr(1,2)}(0) = data_1(0) * data_2(0) \\
&data_{corr(1,2)}(1024) = data_1(1024) * data_2(1024) \\
&for\ i = 1\ to\ 1023: \\
&\quad data_{corr(1,2)}(i) = data_1(i) * data_2(i) + data_1(2048 - i) * data_2(2048 - i) \\
&for\ i = 1025\ to\ 2047: \\
&\quad data_{corr(1,2)}(i) = data_1(i) * data_2(2048 - i) - data_1(2048 - i) * data_2(i)
\end{aligned}
\tag{3.5}
$$

After each of the arrays are low-pass filtered, each one is passed through the <u>findpeaks subroutine</u>, which sharpens the peaks in the arrays. The routine searches for local maxima, and replaces them with impulse functions of the same magnitude. All other points are set to zero. The purpose of this is to simplify the arrays so that the results of the cross-correlation will be relatively clean.

The arrays are <u>cross-correlated</u> in each of the three possible combinations. The FFT method described in section 2.5.3 is used for this process. Each of the three arrays are again FFT'd, the multiplication routine given in equation (3.5) is performed for each of the three possible combinations, and then the three resulting arrays are IFFT'd. Equation (3.5) gives the multiplication routine for the cross-correlation of the data for omnidirectional microphones one and two (the 1-2 combination). The same process is repeated for the 2-3 and 3-1 combinations.

Had ASLAD been implemented with only two omnidirectional microphones, there would have been only one cross-correlation. The range of possible solutions would have been extracted from the cross-correlation result, and the bin containing the largest value would have been identified. That bin number (n) would have been used in equation (2.1) to determine the angular position of the target source. This result would have been passed to the DMPS. The addition of the third omnidirectional microphone complicates the process of determining target position

from the cross-correlation results. The same basic principle is used to determine the target position relative to each pair of omnidirectional microphones, but the results must also be compared in order to obtain a unique solution for the target position. Finally, the result of the sound localizing algorithm must be converted to a solution in terms of the angle ψ for transfer to the DMPS.

Figure 17 is a flowchart showing the steps in the desired position algorithm, used to determine target position from the cross-correlation results. Each cross-correlation result is stored in a 2048 point array. However, there is only a small subset of the 2048 points that represent meaningful results for the sound localizing algorithm. For example, for a 20 kHz sampling rate, there is a maximum possible delay between a sound signal arriving at any two of the omnidirectional microphones of 11 sampling intervals (bins). Therefore, we are only interested in results of the cross-correlation that represent a lead or lag between the signal arriving at two different omnidirectional microphones of 11 sampling intervals or less. The cross-correlation results are stored in wrap-around order, where one half of the range of interest is stored at the end of the array. So, the set of valid solutions if a 20 kHz sampling rate is used can be given by [(2048-n),..., 2047, 0, 1,..., n] where n = 11. Peaks in the cross-correlation result at values of n outside of this range represent time delays between the arrival of a signal at two omnidirectional microphones exceeding 583 microseconds, which cannot be valid solutions to the sound localizing algorithm. For each of the three cross-correlation result arrays, the subset of valid solutions is copied to a new smaller array. For purposes of this explanation, these three smaller arrays will now be termed the *individual solution arrays.* The elements of each individual solution array can be relabelled as [-11,..., -1, 0, 1,...,11]. The size of the individual solution array will vary for different sampling rates. For sampling rates of 15, 22, and 44 kHz, the solution array will span a range [-n,..., 0,...,n] with n = 8, 12 and 25 respectively.

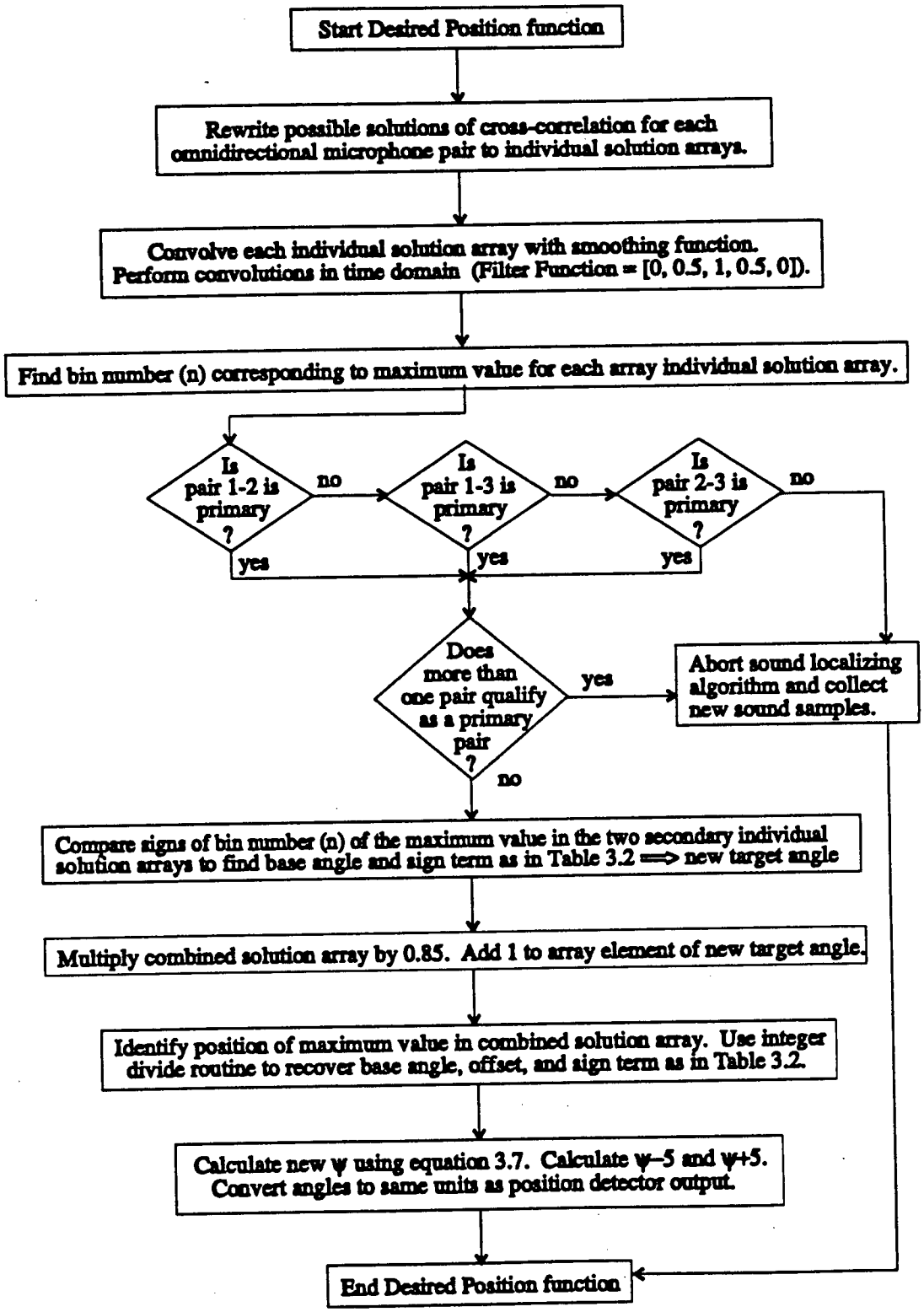


Figure 17: Flowchart of the basic steps in the algorithm used to determine the desired position for the directional microphone (in terms of ψ) from the results of the sound localizing algorithm for each of the three different pairs of omnidirectional microphones.

Each individual solution array is convolved with a smoothing function given by [0, 0.5, 1, 0.5, 0]. The individual solution arrays are small enough that the convolutions can be done efficiently in the time domain. The formula for time-domain convolution was given in chapter 2, equation (2.8). The *peak* for each array is the bin number (n) of the array element holding the maximum value in that array. The position of this peak is the "solution" of the cross-correlation. The peak for each individual solution array is located and recorded.

Figure 8 showed the numbering scheme for the three omnidirectional microphones, the definition of the azimuthal angle ψ, and the ranges for each omnidirectional microphone pair that result in the highest overall resolution for ASLAD. Recall that each omnidirectional microphone pair has two ranges of 60° each, one the mirror image of the other in a plane connecting the two microphones. Only the outward facing range is shown for each omnidirectional microphone pair in Figure 8 to simplify the drawing. To determine which of these pairs of ranges a target source lies in, the subset of each individual solution array representing the 60° range given by ($-30° < \phi_{x\text{-}y} < 30°$) is considered. For the 20 kHz sampling rate, this *60° subset range* is given by the set [-5,...,-1, 0, 1,...,5]. For the 15, 22, and 44 kHz sampling rates, the 60° subset range is given by [-n,..., 0,..., n], with n = 4, 6, or 12 respectively. Because the six 60° ranges defined in section 3.3.1 do not overlap, the peak should lie within the 60° subset of the individual solution array for only one of the omnidirectional microphone pairs. If the peak of more than one of the pairs, or of none of the pairs, lies within the 60° subset of the individual solution array, then no unique solution can be deduced. In this situation, the sound localizing algorithm is aborted, the data is discarded, and a new set of data is collected.

The case where either no pair, or multiple pairs, has its peak in the 60° subset of the individual solution array could result from a number of different circumstances. One circumstance is the case of a sampling interval in which no sources were present. Another

circumstance is a sampling interval in which multiple sources were present, causing either multiple peaks in the cross-correlation solution or peaks representing an average of the multiple sources. By performing the test of comparing the cross-correlation solution from the three different omnidirectional microphone combinations, the chances of localizing only single dominant sources are increased.

For cycles of the sound localizing algorithm that produce a peak in the 60° subset range for only one omnidirectional microphone pair, that microphone pair is called the *primary microphone pairing*. The other two omnidirectional microphone pairings are called the *secondary pairings*. The results of the cross-correlations from the secondary pairings are used to determine which of the two possible ranges for the primary microphone pairings the target is in. It is the signs of the solutions to the other cross-correlations (whether the peak positive or negative) that are used to determine which side of the primary microphone pairing the target is located on. For each range, there is a different combination of positive and negative peaks for the secondary microphone pairings. Table 3.2 shows the unique combinations of the signs of the peaks for the secondary microphone pairings. Once the primary microphone pair is chosen, this table is used to determine which of the two possible ranges is the correct one. The *base angle* is the midpoint of the range, corresponding to $\phi_{x\text{-}y} = 0°$, but expressed in terms of ψ. The *sign term* is necessary to correct for the fact that for the two possible ranges of the primary microphone pairing, in one range a positive peak represents an angle greater than the base angle, while in the other range, the positive peak represents an angle less than the base angle.

Each time a new solution to the sound localizing algorithm is successfully resolved from the three cross-correlation results, it is not sent directly to the control algorithm of the DMPS. The final solution sent to the control algorithm is determined based on a combination of the new

Table 3.2: Look-up table for determination of target range. Base angle and sign term are used in equation 3.6 to determine target angle. (Note: P = primary microphone pairing).

Primary Microphone Pairing	Sign of cross-correlation result (positive or negative peak) for secondary microphone pairs			Range ψ	Base Angle	Sign Term
	1-2	2-3	3-1			
1-2	P	+	-	$30° < \psi < 90°$	60°	+
	P	-	+	$210° < \psi < 270°$	240°	-
2-3	-	P	+	$150° < \psi < 210°$	180°	+
	+	P	-	$330° < \psi < 359°$ $0° < \psi < 30°$	0°	-
3-1	-	+	P	$90° < \psi < 150°$	120°	-
	+	-	P	$270° < \psi < 330°$	300°	+

solution and of solutions obtained in previous iterations of the sound localizing algorithm. An array, called the *combined solution array*, is used to keep track of the contribution of past solutions to final result of the desired position algorithm. The combined solution array is the equivalent of the arrays representing all six of the 60° subset ranges of the individual solution arrays concatenated together (see equation 3.6). Thus, the combined solution array contains a bin representing each of the possible solutions to the sound localizing algorithm over the full range ($0° < \psi < 359°$). For the example of the 20 kHz sampling rate, the 60° subset range is given by [-5,...,-1, 0, 1,..., 5], which has 11 entries. So the combined solution array size for the 20 kHz sampling rate is 66 bins.

$$\begin{array}{l} \textit{combined sol'n array} = \\ \{[-5,..,0,..,5],[-5,..,0.,,.5],[-5,..,0,..,5],[-5,..,0,..,5],[-5,..,0,..,5],[-5,..,0,..,5]\} \\ \quad\uparrow \qquad\qquad \uparrow \qquad\qquad \uparrow \qquad\qquad \uparrow \qquad\qquad \uparrow \qquad\qquad \uparrow \\ \quad \textit{base} \quad \textit{base} \quad \textit{base} \quad \textit{base} \quad \textit{base} \quad \textit{base} \\ \quad \textit{angle} \quad \textit{angle} \quad \textit{angle} \quad \textit{angle} \quad \textit{angle} \quad \textit{angle} \\ \quad \psi=0° \quad \psi=60° \quad \psi=120° \quad \psi=180° \quad \psi=240° \quad \psi=300° \end{array} \tag{3.6}$$

Upon the startup of ASLAD, all entries in the combined solution array are set to zero.

The contribution of past solutions to the result of the desired position algorithm are determined using a weighting function. Each time a new solution is obtained, each element the combined solution array is multiplied by a weighting factor. This weighting factor has a value between 0 and 1, and is a constant value over all elements in the combined solution array (eg. 0.85 × combined solution array). Once each element in the combined solution array has been multiplied by the weighting factor, the entry in the array element corresponding to the new solution to the sound localizing algorithm (as determined by the base angle and peak position relative to that base angle) is then incremented by 1.

The closer the weighting factor is to unity, the slower the decay of past entries, and hence the greater the influence of past results on the determination of desired position. This weighted

average adds stability to the solution, as the sound localizing algorithm must produce the same result for a number of iterations before the directional microphone position will change. It reduces the effect of transient noises on the motion of the directional microphone. It should be noted that the use of a weighing function was not originally considered in the design of ASLAD. It was added because it was found that without it the directional microphone was too "jittery", and in almost constant motion. In addition to being visually distracting, motion of the directional microphone due to the localization of a transient noise would result in the temporary loss of a desired signal. This step does add a delay to the time required to locate and move the directional microphone to a new target. In practice, a value of about 0.85 for the weighting factor provides suitable stability for the microphone without adding too great a delay. The weighting factor of 0.85 results in a delay of 5 cycles of data acquisition and execution of the sound localizing algorithm. Weighting values closer to unity resulted in excessively long delays in the response time for the localization of new targets. Expressed in terms of cycles of data acquisition and execution of the sound localizing algorithm, weighting values of 0.8, 0.75, and 0.7 result in delays of 4, 3, and 2 cycles respectively. A delay of at least 3 cycles was found to be necessary to show some reduction of the jittery motion of the directional microphone.

After adding the new result to the combined solution array, the position of the largest value in the array is determined. This position in the array represents the final result of the desired position algorithm. From the position of the largest value, the base angle and bin number of the largest value relative to that base angle is recovered using an integer divide routine. Referring back to the example in equation (3.6), each of the subsets comprising the combined solution array is 11 elements long. By dividing the position number of the largest value by 11 using integer division, two pieces of information are provided: a quotient and a remainder. Both quotient and remainder are positive integers. The quotient can have six possible values,

numbering from 0 and 5. Each value corresponds to one of the subsets of the combined solution array. A table is used to convert the quotient to one of the six base angles. The remainder can have a value between 0 and 10, and represents the position of the largest value within the subset determined by the quotient. The remainder is corrected by an offset value so that it is centered about the position of the base angle (in this example, the bin number of largest value, relative to base angle bin, would be equal to the remainder term minus 5). Thus, the corrected remainder gives the peak relative to the base angle. The appropriate sign term for the range is then determined using a look-up table much like Table 3.2. Once this information is determined, the final, absolute, angle (ψ) is given by:

$$\psi = \textit{base angle} - (\textit{sign term}) * \arcsin\left(\frac{nc\Delta}{d}\right) \quad \textit{degrees} \tag{3.7}$$

where n is the bin number of the peak (this could be positive or negative), c is the speed of sound in air, Δ is the sampling interval, and d is the separation between the omnidirectional microphones.

Finally, this angle is converted to the same format as the data from the position detector, so that the two may be compared. This is a 12-bit signed integer. The conversion from degrees to "position units" is done using an equation obtained by calibration of the position detector.

3.5.3 Control Algorithm

During the construction of ASLAD, an effort was made to minimize the size and weight of the rotating platform assembly (the directional microphone and platform on which it is mounted). Besides having the benefits of increasing robustness and reducing visual distraction, the reduced inertia of the platform assembly also allowed for greater flexibility in the design of the DMPS control algorithm.

The time required to collect a sound sample for data input to the sound localizing algorithm is long (100 msec for 20 kHz sampling rate). The time required to complete one iteration of the sound localizing algorithm is also long (77 msec). As will be explained further in section 4.1.3, the collection of data during the execution of the sound localizing algorithm was found to result in significant increases to the processing time for the sound localizing algorithm. In order to minimize the additional time required to execute the control algorithm, a simple algorithm, requiring minimal calculation, was sought. The idea was to have a simple control algorithm that could be executed frequently instead of a more complicated algorithm that required tracking of position and velocity data for the purposes of calculating a table of control signal values to be sent to the motor at regular intervals. By not requiring input and output of data at regular timed intervals, the control algorithm can be executed in between the various steps of the sound localizing algorithm.

The control algorithm was implemented entirely in software so that it could easily be altered in the development of the design. Three control methods were tried. They were: an on/off scheme using a single controlling voltage, a proportional scheme where the controlling voltage was varied depending on the distance to be travelled, and a scheme using one of the previous two methods combined with derivative information. The third method was quickly abandoned due to the difficulty in combining the sound localizing algorithm and control algorithm efficiently without increase in processing time, and because the first two methods were both found to perform to the desired specifications without affecting the processing of the sound localizing algorithm.

Four factors were central to the success of the on/off and proportional control schemes. The first is that the efforts made during construction to minimize the weight and inertia of the rotating platform and directional microphone assembly resulted in faster response times on both

startup and change of direction. The second is that the characteristics of the motor are such that for a given driving voltage, the velocity of the motor is constant. Combined with the first factor, this means that the motor can quickly get up to a steady speed, and for all but the shortest distances travelled, the platform assembly is rotating at the same speed as it approaches the desired position. Hence, the shutdown time is roughly the same for all cases. The third factor is that the D/A converter is such that any value written to it is maintained at the output until changed via software. A steady stream of signals to the motor were not required in order to achieve continuous motion of the platform assembly. The fourth factor in the success of the simple on/off and proportional control schemes was the use of the allowed error range and allowed overshoot in the control algorithm. For the directional microphone used, the error range of ±5° allowed would have no effect on the gain of a correctly localized target source. In addition, reduction of the error range would result in either increased oscillations about the desired position before settling, or would necessitate a more complicated control scheme to achieve the higher accuracy. The former is undesirable because the oscillations of the directional microphone about the desired position are visually distracting, and the latter is undesirable due to the increased processing time required and the difficulty of integrating the control algorithm with the sound localizing algorithm.

The result obtained by the sound localizing algorithm is used as one input to the control algorithm for the DMPS. This is the angle (ψ) of the localized target, and is referred to as the *desired position.* The other input is the feedback control signal, which is the angular position of the platform assembly. This is referred to as the *existing position.*

For the on/off control scheme, a negative voltage of a fixed value is applied to the motor circuitry for motion in one direction, and a positive voltage applied for motion in the other direction. The value of the voltage applied was determined by experiment. The minimum

voltage required to overcome the force of static friction and start the platform assembly moving from a stationary position was called the *base voltage*. This was determined by systematically increasing the control signal until the platform started from a stationary position, and also did not stall at any time during rotation. For each increase in the control signal during the experiment, the control signal was set in software to the new value, and presented as a step function from zero, not as an increase over the previous control signal level. The base voltage was found independently for rotation in both clockwise and counterclockwise directions to ensure that any differences in frictional forces or motor operation for the two directions could be identified and compensated for.

Once the base voltage required for dependable operation was determined, the next step was to balance the speed of rotation with the ability to bring the platform assembly to a stop efficiently. It was found that the base voltage was sufficient to allow rotation of the platform assembly through its full range of travel in less than the allowable time of one second, without exceeding the overshoot allowance. In addition, there was a maximum of one overshoot with no subsequent oscillations before coming to rest within the allowed error. This was possible without the need for any special start-up and shutdown method such as using pulses larger than the operating voltage to start and stop the platform assembly as described in section 2.6.4. This made it unnecessary to collect any data on the motion of the platform other than the existing position.

Figure 18 is a flowchart of the control routine. This flowchart applies to both the on/off and proportional control schemes, as the on/off control scheme is a special case of the proportional control scheme. The first step is the calculation of the difference between desired and existing positions. If the magnitude of the difference between these two values is less than 5°, then the directional microphone is correctly positioned within the acceptable limits, and a

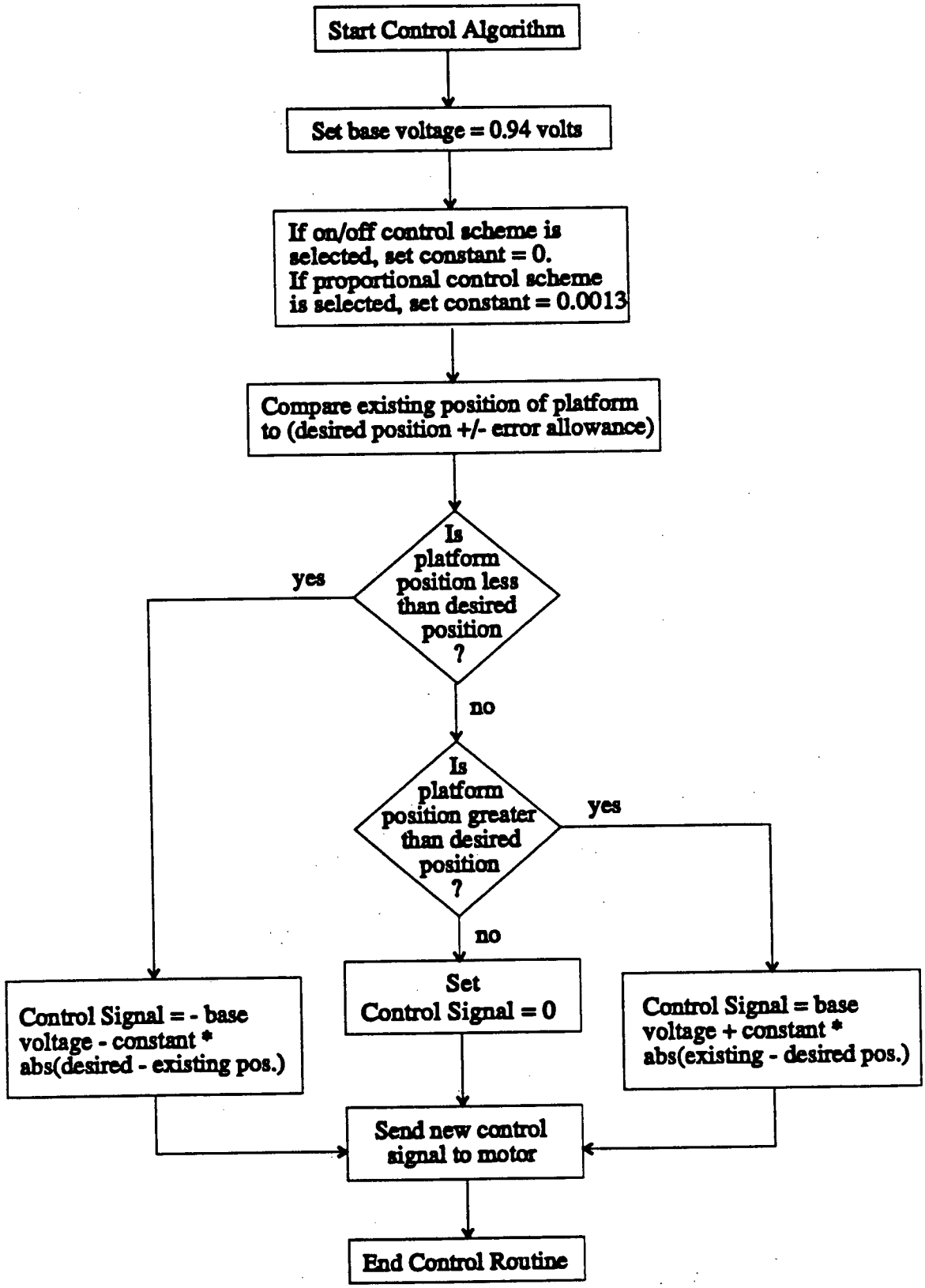


Figure 18: Flowchart of the basic steps in the directional microphone positioning system control algorithm.

control signal of value 0 is sent to the motor. If the existing position is more than 5° lower than the desired position, then a control signal is sent to the motor to rotate the platform assembly in the direction of increasing ψ. If the existing position is more than 5° higher than the desired position, then a control signal is sent to the motor to rotate the platform assembly in the direction of decreasing ψ.

For positioning of the directional microphone, the control scheme resulted in the following events. The control signal is applied to the motor circuitry until the directional microphone is within the error range of the desired position. At that point, the control signal is set to zero, and the platform assembly begins to slow down. In the event that the platform assembly travels through the allowed range without coming to a stop, the control signal for travel in the opposite direction is applied once the directional microphone overshoots the desired position by more than 5°. This signal quickly results in a reversal of direction of the platform assembly, and once back in the allowed range, the control signal is again set to zero. During the direction reversal, the platform assembly does not have an opportunity to gain any speed, and no further oscillations about the desired position occur. For rotation through small angles, there is generally no overshoot. For rotation through larger angles, where the platform assembly is able to reach its terminal velocity, there is generally an overshoot. This on/off control scheme is simple, requires minimal processing time, and can be conveniently inserted in between the various subroutines of the sound localizing algorithm without extending the processing time required for the sound localizing algorithm.

For the case of proportional control, the voltage applied can be given by the following general equation:

$$\begin{aligned} \Delta position &= abs(existing position - desired position) \\ control\ signal &= \pm\ base\ voltage \pm constant * \Delta position \end{aligned} \tag{3.8}$$

where the positive value would result in motion in one direction, and the negative value would result in motion in the other direction. The on/off control scheme is a special case of the proportional control scheme with the constant in equation (3.8) set to zero. The base voltage in equation (3.8) is the same for the proportional control scheme as the base voltage used in the on/off scheme because as the difference between the desired and existing positions approaches zero, this minimum voltage is still required to give reliable operation without stalling. The constant for the proportional term was experimented with in much the same fashion as the base voltage for the on/off control scheme, by slowly increasing the value of the constant in equation (3.8) in the software, and noting the change in operation. The speed of rotation of the platform assembly had to be balanced against the amount of overshoot incurred at higher speeds. The constant of proportionality was chosen such that there was a benefit of reduced time to travel through larger angles with little or no increase in overshoot. The increase in the control voltage due to the proportional term was limited to a maximum of 30% of the base voltage. Beyond this increase, the overshoot allowance was exceeded. Very little change in the performance of the control algorithm was noticed for rotation through small angles, but the time required to travel through larger angles was reduced.

Examples of the performance of the on/off and proportional control schemes are given in chapter 4. Both met the performance criteria discussed in section 3.2.4. The on/off control scheme was used in the assessment of the specifications in section 4.1.4, and was also used for the testing of ASLAD in an anechoic chamber (section 4.2). The only difference between the two control schemes reflected in the specifications is the time required for travel through the full range of travel.

CHAPTER 4: Experiments and Results

4.1 ASLAD Specifications

In chapter three, a table of preliminary design specifications for ASLAD was presented. In this section, the specifications for the final design are given. Table 4.1 is a repeat of table as 3.1, with a new column of the values, selected or measured, for ASLAD. Any changes made to the preliminary specifications (Table 3.1) are described in the following sections, along with a more detailed report on the performance of each of the different stages of the devices.

4.1.1 Stage 1 -- Input stage

The initial design value for the sampling rate was 44.1 kHz. As explained in section 3.5.1, it was necessary to lower the sampling rate in order to minimize processing time. To achieve the full benefit of the fast processing times of the DSP chip, the arrays being FFT'd and inverse FFT'd had to be loaded into the 2 kilobytes of on-chip RAM. At the same time, a relatively long sound sample, approximately 100 millisecond duration, was desired. By reducing the sampling rate to 20 or 22 kHz, a 2000 sample array of 100 or 90.9 milliseconds in length (respectively) could be collected, while keeping the processing time to a minimum. The resolution of thesound localizing algorithm was still under ±6° for these lower sampling rates. In Table 4.1, the "ASLAD VALUE" column is split into two sections for the parameters affected by a change in sampling rate, and results for both 20 kHz and 22 kHz sampling rates are listed.

In section 3.4, the benefits of using a low-pass filter already built into the A/D converter were outlined. The filter is a third order Butterworth filter, custom tuned with plug-in resistor packs on the A/D board. The correct value of the resistor packs required to obtain a 10 kHz roll off could not be located at the time of construction. The closest available value was substituted,

Table 4.1: Specifications sheet for ASLAD.

STAGE	PARAMETER	PREFERRED VALUE	ACCEPTABLE LIMIT	ASLAD VALUE	
1. Input stage	Sampling rate	44.1 kHz		20 kHz	22 kHz
	Filter Bandwidth	10 kHz		7.25 kHz	
	Voltage range		±2.5 volts	±2.36 volts	
	Output impedance of omnidirectional microphones			1 kΩ	
	Omnidirectional microphone separation	20 cm		20 cm	
2. Sound Localizing Algorithm	Resolution			±5.8 degrees (worst case)	±5.4 degrees (worst case)
	Processing time			77 msec	
3. Data transfer	Time between updates	100 msec		177 msec	168 msec
	Data format			12 bit signed integer representing target angle (ψ)	
4. Directional Microphone Positioning System (DMPS)	Total range (x) of system	270 degrees	180 degrees	240 degrees	
	Time allowed for x degrees of travel	1 second		0.9 seconds	
	Overshoot	±5 degrees	±10 degrees	±10 degrees	
	Settling time	100 msec		90 msec	
	Travelling speed (max.)	270 degrees/second	180 degrees/second	270 degrees/second	
	Tracking speed			15 degrees/second	
	Error on feedback signal	±2.5 degrees	±5 degrees	±5 degrees	
	Mean positioning error			±2.5 degrees, std. dev. = 1.9	
	Max. positioning error			±5 degrees	

giving a low-pass filter with a bandwidth of 7.25 kHz. A lower frequency as opposed to a higher one was chosen for the -3 dB point because the sampling rate had also been changed during the course of the design process. With this filter bandwidth, both the 20 and 22 kHz sampling rates can be used without fear of encountering aliasing problems.

The limits of the voltage range for the input stage are well within the ±2.5 volts range of the A/D board. The signals from the omnidirectional microphones are limited to ±2.2 volts with hard limiters, and the range of the position feedback signal is from 0 to 2.36 volts. The output impedance of the omnidirectional microphones used is 1 kΩ. Other microphones could be used if desired, with little change to the circuitry. A buffer amplifier is used in the input stage to bypass the effects of a parallel 10 kΩ resistor in the A/D circuitry, and so only an adjustment of the gain would be necessary if there was a need to replace the existing microphones.

The three omnidirectional microphones are arranged at the vertices of an equilateral triangle of side 20 cm. Thus the omnidirectional microphone separation for each pair is 20 cm. The omnidirectional microphone separation is fixed, and could not be altered without repackaging the base of ASLAD.

4.1.2 Stage 2 -- Sound Localizing Algorithm

The resolution of thesound localizing algorithm varies with both sampling rate and angle of incidence on a given pair of microphones. By using three microphones in a triangular arrangement, as described in section 3.2.2, and a sampling rate of 20 kHz, the worst case resolution due to the finite sampling rate is limited to ±5.8°. The best case resolution is ±5°. The best and worst case resolution for a 22 kHz sampling rate is ±4.5° and ±5.4° respectively.

The processing time required to execute one cycle of thesound localizing algorithm for the three omnidirectional microphone configuration is 77 milliseconds. When compared to the

time of 44 milliseconds required to process the sound localizing algorithm for a single pair of microphones, this is a 75% increase in processing time. This is very reasonable in terms of the benefits gained. In particular, the error in the localization is reduced drastically, the range over which localizations can accurately be made is increased to a full 360°, the ambiguity over which side of the microphone pair a sound might be coming from is eliminated, and the mechanical construction is considerably more robust if not more elegant.

4.1.3 Stage 3 -- Data Transfer - Sound Localizing Algorithm to Directional Microphone Positioning System

In section 3.2.3 the rationale for choosing a time-between-updates of 100 milliseconds was outlined. However, with the addition of the third omnidirectional microphone, and the increase processing time that resulted, it became necessary to change this value. The original design was based on the idea that the signal processing required to determine the target position from one set of data would be done in parallel with the collection of the next sound sample. To test this method, a version of the software was written for a two omnidirectional microphone configuration in which thesound localizing algorithm was executed for one set of data in parallel with the collection of a new set of data. With a 15 kHz sampling rate, for which a period of 133.3 milliseconds is required to collect a 2000 point sample, and with an on/off control scheme incorporated into the interrupt routine, it took 104 milliseconds to complete the processing that took only 44 milliseconds when uninterrupted. This means that beyond about a 17 kHz sampling rate, the processing time would be larger than the data collection period, and the benefits gained by collecting and processing the data in parallel are lost. For the three omnidirectional microphone configuration, using the sampling rates desired, it follows that it is more efficient to first collect the data, and then process it without interruption. For a sampling rate of 20 kHz,

the total time for one cycle is 177 milliseconds (100 for data collection and 77 for processing). For a 22 kHz sampling rate, the total time for one cycle is 168 milliseconds.

The output of thesound localizing algorithm is the angle (ψ) of the identified target. The angle ψ is defined in Figure 8. In addition to ψ, the values (ψ - 5°) and (ψ + 5°) are also calculated. All three values are then converted to the same units as that of the position feedback signal to allow a direct comparison. The data from the position detector is left as a twelve bit signed integer, which is the output form for the A/D converters used. The twelve bit signed integer corresponds to the voltage across the leads of the position detector.

4.1.4 Stage 4 -- Directional Microphone Positioning System

The total range of the directional microphone is 240°, which is well within the range set out in the design specifications.

A maximum time allowed for the platform to move through this range of travel was chosen as one second in the design specifications. The time required for travel through the full 240° was measured at approximately 0.9 seconds. This is a conservative measure of the parameter, as it was measured using an on/off control scheme. The time required for travel through 240° is as low as 0.7 seconds if the proportional control scheme is used. Further reduction in the travel time using the on/off and proportional control schemes cannot be made without exceeding the allowable overshoot or causing multiple oscillations about the desired position before settling.

The overshoot is defined in absolute terms, and is set in the software. It can easily be changed. The value for overshoot is related to the speed with which the platform assembly rotates. The faster the rotation velocity, the higher the overshoot. The simplest control scheme is obtained by balancing the rotation velocity and the allowed overshoot such that the platform

is able to settle to a stationary position with minimum oscillation about the desired value. For both the on/off and proportional control schemes a control signal or formula determining the control signal was found through experiment such that the maximum overshoot was 10° over the desired position, or 5° over the allowable error range on the desired position. In addition, other than one initial overshoot peak, there were never any further oscillations about the desired position. The maximum overshoot is only experienced for cases where the directional microphone is being moved through most of its range of travel (eg. 180° or more). For rotation through smaller angles there is less, and often no, overshoot (Figure 19).

A settling time of 100 milliseconds was chosen for the preliminary specifications. This value was set to match the time-between-updates specification. This settling time is the velocity settling time for the platform, and if it was higher than the time between updates, then the system would not be able to accurately follow changes in the control signal. The time-between-updates was increased with the addition of a third omnidirectional microphone, and as a result, the settling time specification could have been relaxed. However, the settling time for the platform is lower than the initially specified time, at about 90 milliseconds. The settling time was measured by applying the control signal to the directional microphone platform assembly at rest. This is the longest settling time, as it includes the time required to overcome static frictional forces.

The travelling speed is equivalent to the total range of system divided by the time required to travel through the total range. Based on the values in table 4.1, this is calculated as 270°/second.

Continuous tracking of a moving sound source was not a goal for ASLAD, and so there is no software designed to perform such a task. However, ASLAD's ability to track a moving target can be estimated using the time required to locate a new sound source, the acceptance

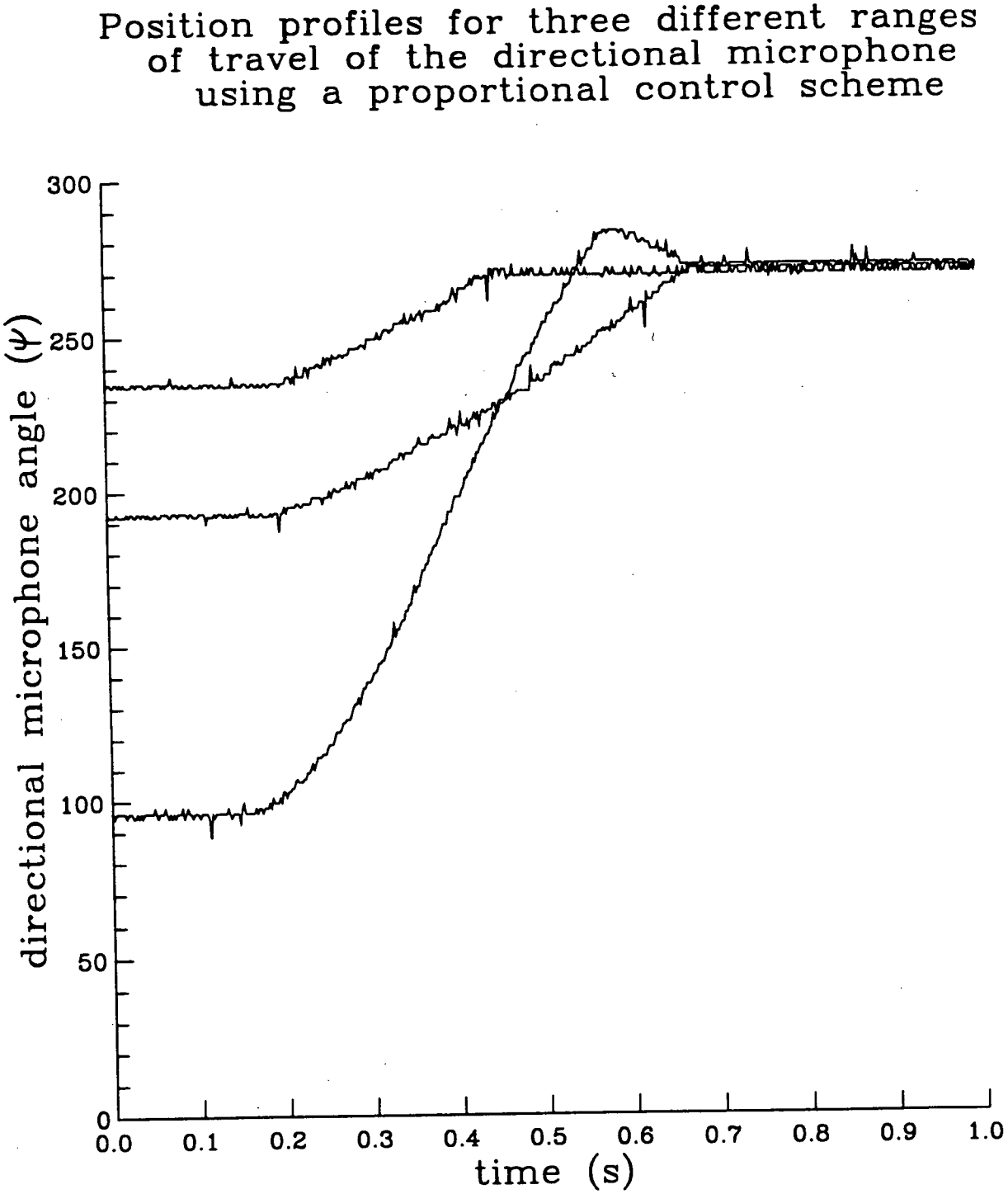


Figure 19: Example of position profiles for three different ranges of travel using a proportional control scheme. The graph shows that the time required to travel through larger angles is reduced with the proportional control scheme, as the travel times for rotation through 75° and 175° are almost equivalent. However, there is overshoot for the case of travel through the larger angles.

angle of the directional microphone, and the time required to move through a distance equal to the acceptance angle of the directional microphone. This is an estimate based on the idea of "piecewise" tracking of a source. That is, because the acceptance angle of the directional microphone is not a pencil beam, but is about 60°, a continuously moving target could be tracked by moving the platform in small steps, without the target source ever leaving the acceptance cone of the directional microphone. Take 30° as the size of the step to be taken with each move. From a stationary position, it takes about 150 milliseconds to move 30°. As explained in section 3.5.2, the time delay for a new target to be registered must be at least 3, and preferably 5, cycles of data acquisition and execution of thesound localizing algorithm in order for the system to be less sensitive to transient noise. Also, as was determined in a set of tests performed in an anechoic chamber (section 4.2.4), for voice an average of 6 cycles of thesound localizing algorithm were required to localize a source. If these two delays are added together, it can be estimated that 11 cycles of thesound localizing algorithm would be required to determine that the source had moved. Using this estimate, about 1.95 seconds are required to move through the 30°. Normalizing to degrees per second, an estimate of 15° per second is obtained for ASLAD's maximum tracking rate.

The <u>accuracy</u> of the DMPS was tested by bypassing thesound localizing algorithm in the software, presetting different values for the desired position, running the program, and measuring the position error. For this experiment, an equal number of trials were done for rotation in both clockwise and counterclockwise directions. The desired positions were chosen such changes in angle ($\Delta\psi$) ranging from 5° through 240° were tested. The error in reading the position of the directional microphone was estimated at ±0.5°. The mean error in positioning was ±2.6°, the standard deviation on the error was 1.5°, and the maximum error was ±5°. The sample size was 46. For one direction there was a slight increase in the average error above a travel of 60° (mean

error for measurements above 60° was 1.3° higher than for below 60°), but for the other direction the results were consistent for all ranges of travel. This difference in the behaviour of the DMPS for the two directions of travel is most likely due to the fact that the contact brush for the potentiometer (position feedback) is asymmetrical, and is in one direction being pulled and in the other direction being pushed across the resistive element. There would be slight differences in the frictional components for each case.

Some examples of the control signal and the position data are shown in Figures 20(a) and (b). Figure 20(a) shows an on/off control scheme. Figure 20(b) shows a proportional scheme. Due to the time required to execute thesound localizing algorithm, the data collection interrupt routine and the signal processing could not be done simultaneously, as it would result in very high cycle times. In the three omnidirectional microphone system, the most consistent way of applying the control algorithm was to invoke it periodically. For a 20 kHz sampling rate, an update of the desired position variable is produced every 177 milliseconds. In that time the control algorithm is invoked 35 times. The calls to the control routine are made on average every 5 milliseconds, and are either executed between calls to other subroutines, or during the collection of sound samples. The rotation of the platform between calls to the control routine is at most 5° when travelling at full velocity. As can be seen from Figure 20(b), the calls to the control routine are made often enough that even though the change in the control signal is occurring in steps, the plot of the control signal appears to be smooth.

4.2 Performance Evaluation in an Anechoic Environment

In order to characterize the performance of ASLAD as a whole, a set of tests were performed in an anechoic chamber. In these tests, the ability of the device to localize sounds under ideal conditions, the signal-to-noise ratio required for normal operation, the time required

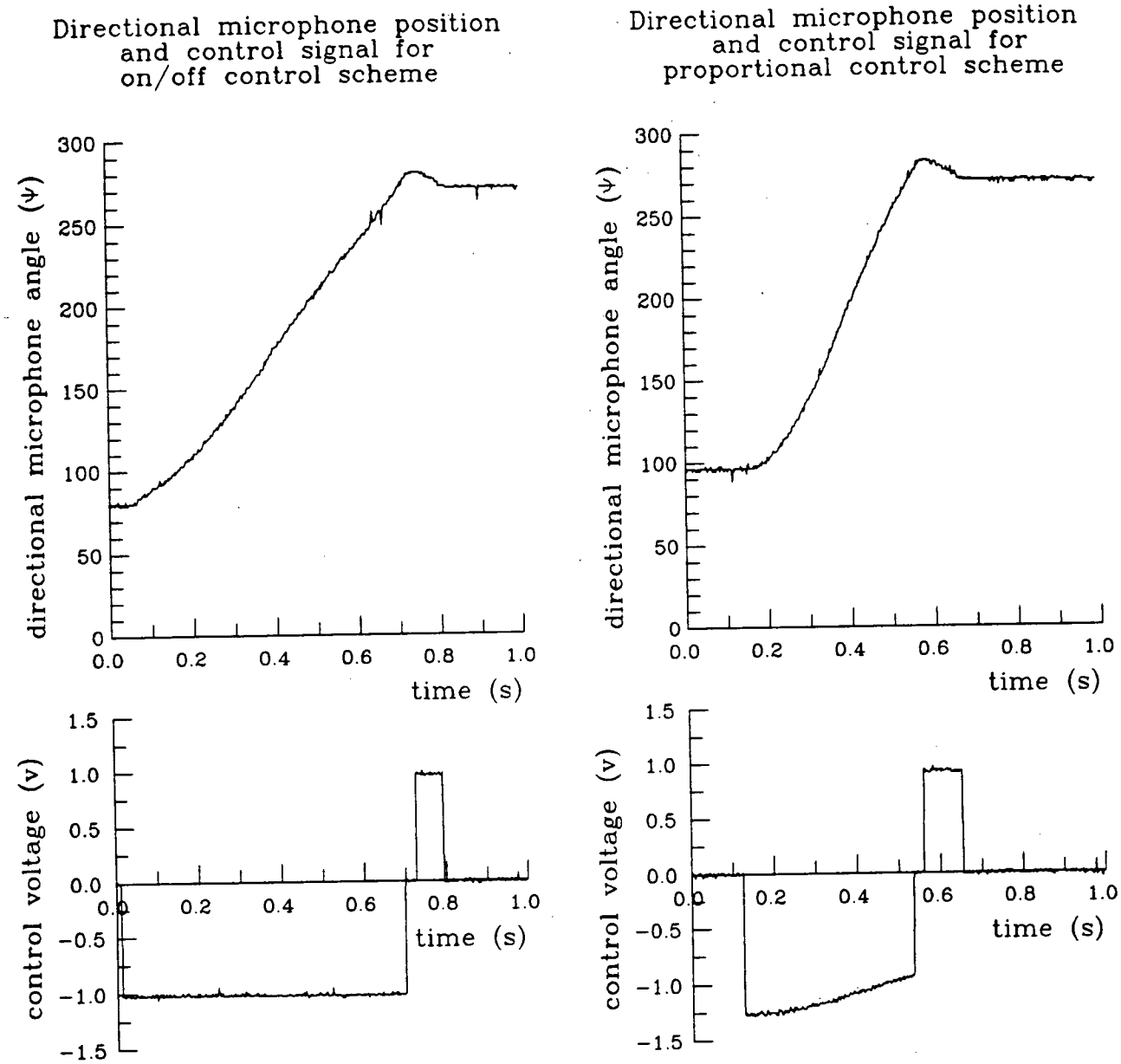


Figure 20: Position profiles and the corresponding control voltages for (a) on/off and (b) proportional control schemes.

to correctly identify and move to a sound source, and the behaviour of the device when subjected to two sources simultaneously, were studied systematically. Figure 21 is a diagram of the experimental setup in the anechoic chamber. The chamber floor has dimensions of approximately four meters by four meters. The chamber has one rotating boom mounted in the centre of the ceiling, and a second one mounted in one corner of the ceiling. The floor consists of a woven wire mesh strung above sound absorbing cones. Because this floor is very bouncy, it was necessary to hang a platform from the ceiling for ASLAD to sit on. The device was centered under the axis of rotation of the central boom. Two different types of sounds were used in the experiments: target sounds, which were to be located by ASLAD; and noise sounds, which were used as interference in the determination of the required signal-to-noise ratio for reliable operation. A speaker used to present noise was mounted directly above the device, on the underside of the boom. A speaker used to present target sounds was mounted on the vertical arm of the rotating boom, as shown in Figure 21. The position of the target relative to ASLAD was defined by three parameters. Distance is defined as the horizontal separation between the center of ASLAD and the target speaker, height is defined as the height of the target speaker above the top surface of ASLAD, and angle refers to the angle of rotation of the boom on which the target speaker was mounted. This angle of rotation was measured using the same scale (ψ) as that for ASLAD, which was defined in Figure 8. A sheet of polar graph paper was centered on the top surface of ASLAD, and both the angle of the directional microphone and the angle of the boom were measured on this scale. The former was done using a pointer attached under the directional microphone, and the latter was done using a plumbob suspended from the boom such that it hung directly over the scale. The following sections contain more specific information regarding the setup for each experiment, and the results of the experiments performed.

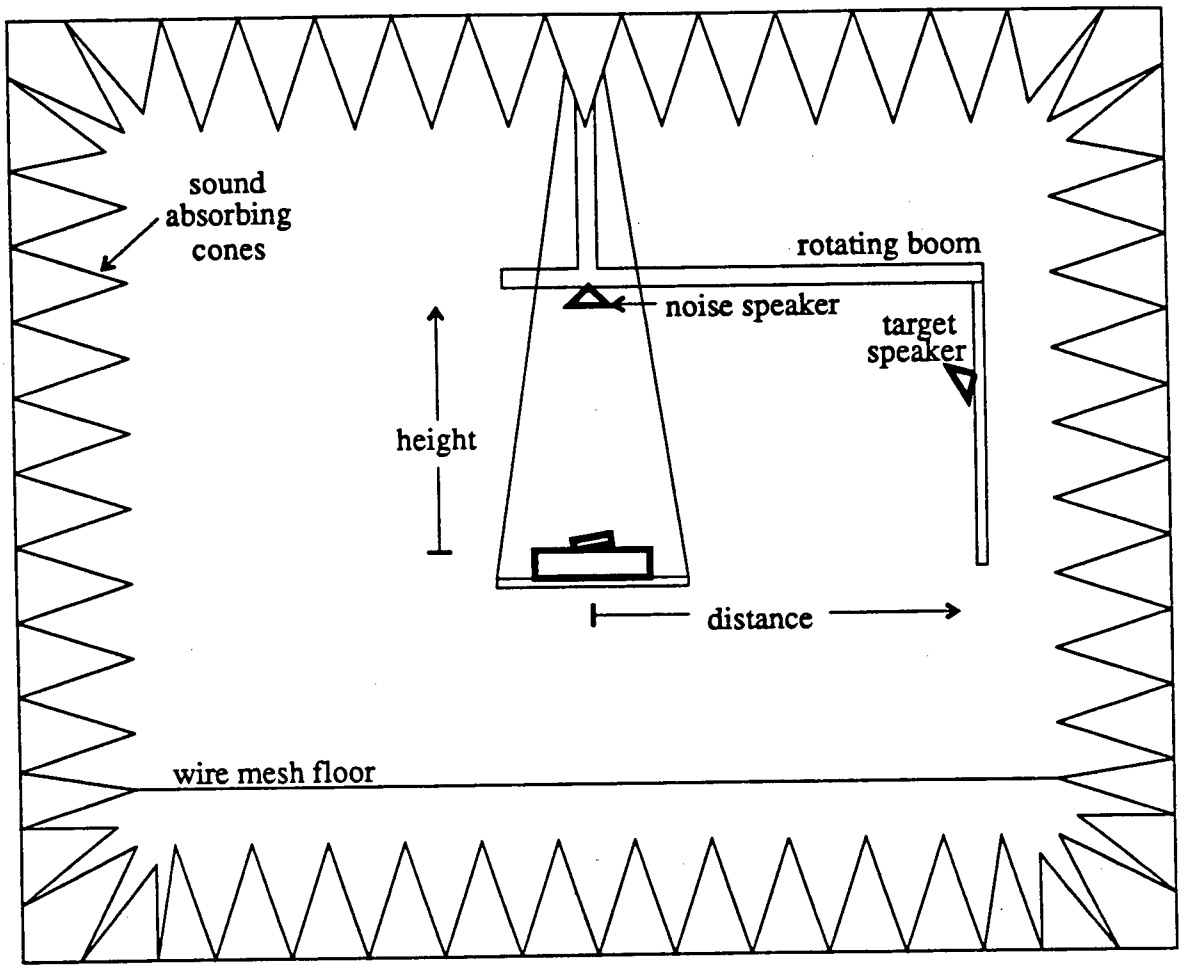


Figure 21: Diagram of the test setup in the anechoic chamber. ASLAD was positioned on a platform suspended from the ceiling. The target speaker, used to present sounds to be localized, was mounted on a rotating boom. Target position was measured using the parameters of distance and height as shown on the diagram, and using the angle ψ as defined in Figure 8.

4.2.1 General Observations

All measurements of sound pressure level[2] (SPL) were made using a Rion NA29E sound level meter. The average SPL measured at the position of the device with no sound sources present varied between 50 and 55 dB, measured using a flat weighting scale across the frequency range from 20 to 12,500 Hz. Much of the energy in this background noise was in the two lowest octave bands, having center frequencies of 31.5 and 63 Hz. It is most likely that the source of this low frequency noise is the mechanical coupling of the room, via the concrete floor on which it is built, to the heating and ventilation systems of the building. Other machinery in the building may have also contributed to the background noise.

The experiments were performed using a 20 kHz sampling rate for the collection of data from the omnidirectional microphones, and the on/off control scheme in the DMPS.

In order to select target sounds that could be used for the experiments, some initial investigations were made into the frequency range that the device would respond to, and also into the SPL required for consistent results. At a distance of 12 inches from, and height of 0 inches above, ASLAD, a sinusoidal signal from a function generator was presented at 100 Hz intervals. It was found that above 1100 Hz the ability of the performance of the device deteriorated rapidly, displaying erratic behaviour and the inability to settle at any one location. Below 1100 Hz, however, there was no variation in the ability of the device to localize over the frequency range. Step by step simulations of thesound localizing algorithm using the commercial software package MATLAB confirmed that this should be the case. For a 20 kHz sampling rate, the maximum number of sampling intervals (bins) between the arrival of a signal at two of the omnidirectional

---

[2] Sound pressure level is defined as follows:

$SPL = 20 \log_{10}(P_{sound} / P_{ref})$ dB, where P_{sound} is the pressure of the sound being measured, and P_{ref} is a reference pressure of 20 micropascals.

microphones is 11, which is obtained for a target at position **C** or **C'** in figure 4(a). The first step of thesound localizing algorithm is to rectify the signal, so for a sine wave target, the signal repeats itself every half-wavelength. Consequently, we can only consider one half of a wavelength as being unique. If the half-wavelength is smaller than the separation between the two microphones (or equivalently, if the half-period of the sine wave is less than 583 microseconds), then the cross-correlation will yield more than one peak, representing more than one possible solution. For a 20 cm separation between the omnidirectional microphones, the highest frequency sine wave signal that will give, over all possible angles, only one solution to thesound localizing algorithm per microphone pair is 857 Hz. Once the target signal exceeds this limit, the combination of the results from the three omnidirectional microphone pairs will no longer fit any of the possible solutions laid out in the desired position function of thesound localizing algorithm. The results are either discarded based on the inconsistencies, or may be incorrect if a single solution is obtained. ASLAD was capable of functioning for a small range of frequencies above this 857 Hz limit. For a small range of frequencies above the 857 Hz limit, there is a subset of the range of ψ where unique solutions can be obtained. However, as the frequency increases, this subset decreases, until a point is reached where no unique solution exists for targets at any angle ψ.

The SPL required to give consistent results was determined by starting with a signal having an SPL of approximately 60 dB, and increasing the SPL of the signal in steps of approximately 3 dB at a time. The performance of the device was noted at each level. This was done for a variety of frequencies. On average, the device was able to localize approximately 50% of the time at 65 dB SPL, 75% of the time at 70 dB SPL, and 90% of the time or better at 75 dB SPL. In addition, the localizing ability for some signals began to deteriorate again above 85 dB SPL. This could be due to either the speaker being overdriven, causing distortion,

or by non-linearities in the signal caused by the hard-limiters in the circuitry for the omnidirectional microphones. As a result of the above information, the signals used for the remainder of the experiments were kept within the range of 75 to 85 dB SPL wherever possible.

Figure 21 shows a vertical bar attached to the end of the rotating boom. The target speaker was mounted to this bar for the localization experiments. The height of the target speaker (as defined in Figure 21) was adjusted by clamping it at different positions on the vertical bar. There was, unfortunately, no convenient way of adjusting the distance from ASLAD to the target speaker. The vertical bar (onto which the target speaker was mounted) was not adjustable in any way (i.e. there was no way to attach it to the rotating boom at any point other than the end, as shown in Figure 21). The anechoic chamber was only available for a very short period of time, so it was not possible to construct a custom mounting system for the target speaker. The distance between the center of ASLAD and the target speaker mounted on the vertical bar was 48 inches (122 cm). At this distance, it was not possible to obtain a signal in the desired SPL range for low frequencies. Below 400 Hz, it was not possible to achieve SPLs of above 71 dB without overdriving the speaker and causing distortion. At 600 Hz, SPLs up to about 75 dB were possible without distortion. With an 800 Hz signal, it was possible to have SPLs in the desired range of 75 to 85 dB without distortion. In order to present a signal containing both high and low frequency components, a higher frequency sinusoid modulated by a low frequency triangular wave was tried. This type of test signal was found to be a suitable source for localization. An 800 Hz sinusoidal waveform modulated by a triangular wave with a frequency of 5 Hz was chosen as the source to be used for the localization experiments. For the remainder of this chapter, it shall be referred to as the *amplitude modulated (AM) sine signal.* This type of signal has the same basic properties as speech, but is not as complicated, as there is only one frequency being modulated. In speech there are many frequencies being amplitude

modulated by the movement of the tongue and mouth.

4.2.2 Errors in Sound Localization

Two different studies were done to determine the average error in localization of a source. The first study was done with the AM sine signal presented at the target speaker. The target speaker was mounted on the vertical arm of the rotating boom as shown in Figure 21. No noise source was used in this set of experiments. Measurements of the error in sound localization were made over a 120° range ($120° < \psi < 240°$) in intervals of approximately 10°. This was the largest range possible without reconfiguring the entire setup, as below $\psi = 120°$ and above $\psi = 240°$ the rotating boom hit the wires used to hang the platform on which ASLAD was positioned. Reconfiguring the setup would have likely lead to inconsistencies in the data. Measurements were taken for four different heights of the target speaker. Average error and standard deviation for the each of the different heights are shown in table 4.2. The error in measuring the position of the target speaker was estimated at ±3°. The error in measuring the position of the directional microphone was estimated at ±0.5°, which was half the smallest scale division on the polar plot used to measure the angle.

The average error measured in these experiments is the combined localization error of both thesound localizing algorithm and the DMPS. In all cases, the error was not greater than that required for the target to fall within the acceptance angle of the directional microphone. The average error for the case of height = 25 cm was slightly worse than for the other three cases. Of the four different heights used in the experiments, this is the height for which the directional microphone was most likely to shadow one of the omnidirectional microphones. This effect was most predominant at or about the $\psi = 120°$ and $\psi = 240°$ positions. In some cases, there were 180° reversals of the position of the directional microphone when the target speaker was at either

Table 4.2: Mean error and standard deviation for sound localization experiments with AM sine signal.

DISTANCE (±1 cm)	HEIGHT (±1 cm)	SPL (±0.5 dB)	NUMBER OF TRIALS	MEAN |ERROR| (degrees)	STANDARD DEVIATION
122	0	79.2	13	4.2	3.3
122	25	80.0	13	6.4	3.9
122	51	80.9	10	4.9	3.3
122	76	80.5	10	4.8	2.9

Table 4.3: Mean error and standard deviation for sound localization experiments with live voice.

DISTANCE (±2 cm)	HEIGHT (±2 cm)	SPL (±3 dB)	NUMBER OF TRIALS	MEAN |ERROR| (degrees)	STANDARD DEVIATION
30	0	75	25	3.1	3.0
61	25	75	25	3.2	3.3
61	0	75	25	4.4	3.1
91	25	72	25	3.6	2.4
91	0	75	25	3.2	2.7

Table 4.4: Time required to locate and move to a new target.

INITIAL SOURCE	FINAL SOURCE	NUMBER OF TRIALS	MEAN TIME (seconds)	STANDARD DEVIATION
silence	voice	9	1.56	0.43
silence	AM sine	7	2.46	0.96
one of: AM sine or continuous sine	other of: AM sine or continuous sine	20	3.97	1.77

the $\psi = 120°$ or $\psi = 240°$ position, where one of the omnidirectional microphones was shadowed by the directional microphone. In these cases, by standing directly behind ASLAD (opposite the target speaker), the sound field could be perturbed enough to allow an adequate signal to reach the shadowed omnidirectional microphone, thus causing the directional microphone to turn back to the desired position. Furthermore, in other cases when the directional microphone was pointing at the target speaker, locked onto the correct position, the directional microphone would sometimes become jittery and cease to be able to lock onto the correct position if the sound field was perturbed by standing or moving about behind ASLAD. These simple experiments suggest that echoes are a real problem for the system, and incorporating a method for handling them into thesound localizing algorithm will be of great benefit in the performance of the device in non-ideal acoustic conditions.

The second study done to evaluate the ability of the device to localize sound sources was performed using live voice as the test source. Although the use of voice is more difficult to control in an experimental setup such as this, it was thought that because thesound localizing algorithm is based on the characteristics of speech, and the ability to follow a live conversation is the primary goal of this exercise, the results of this study would be very interesting. Data was taken over a 240° range ($60° < \psi < 300°$) in 10° intervals, at five different combinations of distance from and height above ASLAD. Target height was measured from the top surface of ASLAD to mouth level. The target distance and angle (ψ) were measured using a piece of string, marked at the required distances, fixed to the underside of the platform on which ASLAD sat. The string was fixed such that it would pivot about ASLAD's centre point, and the marked distances would scribe imaginary circles. The target angle (ψ) was measured by aligning the taut string with the desired mark on the polar plot used to measure the position of the rotating boom and the directional microphone. The error on this measurement was estimated at ±3°. As before,

the error associated with reading the position of the directional microphone was estimated at ±0.5°.

Voice level was measured in two ways. One way was using the SPL meter that was used for the SPL measurements during the course of these experiments. The meter has a digital readout as well as a digital scale. The display of the meter was mounted such that it could be seen from all locations about ASLAD, and its microphone was suspended over the center of the device. The second method used to monitor voice level was by using a Volume Unit (VU) meter made in our laboratory. The VU meter consisted of an analog meter movement connected to the output of a microphone, amplifier, and rectifier circuit. This device was calibrated using an 800 Hz sinusoidal signal whose SPL was measured with the SPL meter. The scale was adjusted such that the target SPL of 75 dB gave a reading of 0.58 on a scale from 0 to 1. This was marked as the 0 dB point for the VU meter. On this scale, a 5 dB change in SPL resulted in a 0.2 change in the reading of the meter. The test pattern used was repeated counting from one to ten, and with some practice, it was possible to present this pattern with the meter consistently peaking within 3 dB of the required SPL. All tests in this set of experiments was done using voice presented at 75 dB SPL, with the exception of one data set (distance = 91 cm, height = 0 cm), for which the voiced test pattern was presented at an average level of 72 dB SPL. The results for this set of experiments are given in table 4.3.

The ability of ASLAD to localize live voice in the anechoic environment was very good. The average errors were well within the total sum of the allowable errors for each of its different stages. There were, out of the 125 measurements made using voice as the test source, 6 instances where the error was 10° or higher, with the highest error being 14°. Even these larger errors would result in the target being within the acceptance angle of the directional microphone. The system was also stable during the testing with voice. There were no cases of 180° reversals of

directional microphone position. The directional microphone moved to the desired location with little or, as was most often the case, no searching or random motions. If the target signal continued once the directional microphone was pointing at it, the directional microphone did not search around or lose the position. When the target signal stopped, the directional microphone remained in its last position until presented with a new stimulus.

For both of these different target signals, the average error in localization was very good, and in all cases, the average error was much lower than the sum of the allowable errors for each stage of the device. From this, it can be concluded that under ideal conditions, thesound localizing algorithm works within specifications, and the system as a whole is capable of localizing a source and moving the directional microphone with sufficient accuracy to allow the source signal to be picked up by the directional microphone.

4.2.3 Behaviour in Noise

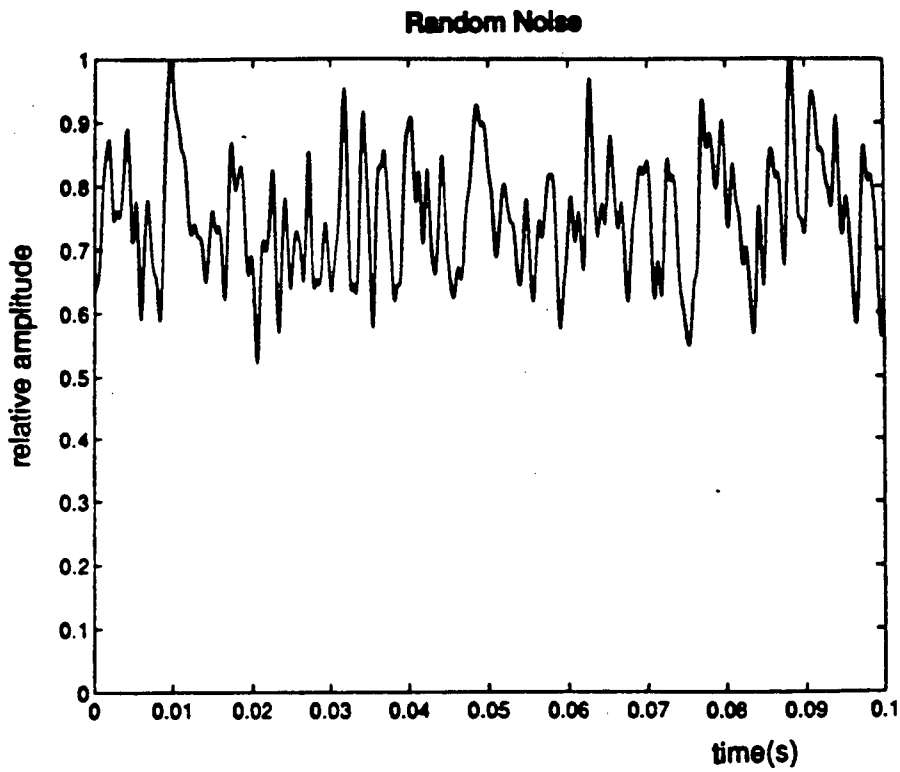
Two different noise sources were used in the experiments to identify the lowest signal-to-noise ratio at which ASLAD will still function. For both experiments, the speaker used to present the noise source was mounted directly above the device, as shown in Figure 21, at a height of 80 cm above the top surface of the device. It was not possible to "flood" the room with noise from multiple speakers, so the best possible setup for these experiments was to present the noise from above the device such that it was uniform over the three omnidirectional microphones. The speaker used to present the target signal was mounted on the rotating boom at a distance of 122 cm from ASLAD and a height of 25 cm above it. All SPL measurements were made at the center of the array of omnidirectional microphones.

The first test was made with white noise. The target signal was the previously defined AM sine source. The background SPL, with no noise or test signal, was 50.7 ± 0.5 dB. The

target signal was set at 70.2 ± 0.5 dB SPL, as the maximum available noise signal was 70 dB SPL. The noise signal was first set at this maximum level, and reduced in steps of approximately 3 dB. The performance of the device was recorded at each step.

The following results were obtained. Below a signal-to-noise ratio of 8 dB SPL, ASLAD could not localize the target, and the directional microphone was in constant, random motion. With a signal-to-noise ratio of 8 ± 1 dB SPL the directional microphone was repeatedly turning to the target angle, stopping briefly, then moving away again in a random fashion. With a signal-to-noise ratio of 11 ± 1 dB SPL, the directional microphone was able to move to and remain at the target angle for longer periods of time, but still exhibited some erratic behaviour, moving away occasionally and then returning back to the target angle. With a signal-to-noise ratio of 15 ± 1 dB SPL, ASLAD was working as well as if there was no noise present. It is interesting to note that 15 dB SPL is a common signal-to-noise ratio for signal comprehension in noise. Individuals with normal hearing require, on average, 10 to 15 dB SPL signal-to-noise ratio for speech intelligibility (Duquesnoy, 1983), and individuals suffering from hearing loss often require anywhere from a 5 to 15 dB increase in the signal-to-noise ratio over those with normal hearing (Soede, 1993a). Automated speech interpreters generally require a high signal-to-noise ratio for good performance, as evidenced by the fact that most will only give good results if the used speaks directly into the microphone used to collect the speech samples.

The white noise signal is a random signal, and is composed of many frequencies. By generating a 2048 point white noise array using a random number generator, rectifying it, filtering it with the same low-pass filter as that used in thesound localizing algorithm of ASLAD, and normalizing it, the waveform in Figure 22(a) was obtained. By comparison, Figure 22(b) is a rectified chirp signal after the application of the same low-pass filter, and normalization. The chirp is also composed of many frequencies, but because they are ordered in a pattern of steadily



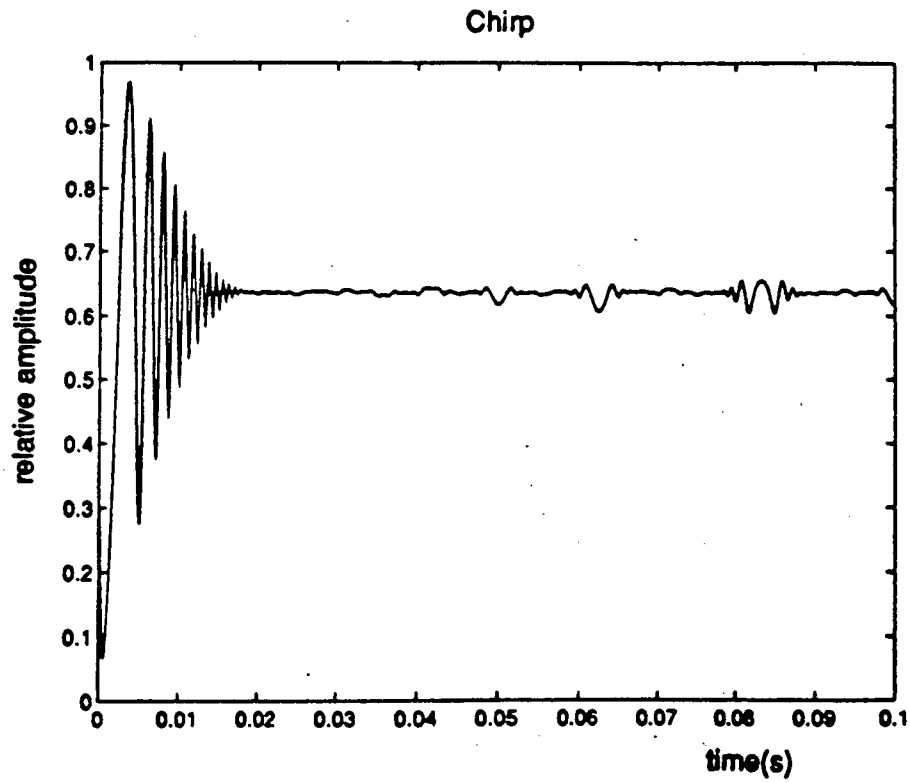


Figure 22: Examples of (a) random noise, and (b) a chirp signal, after rectification, low-pass filtering and normalization. The chirp signal has unique characteristics, suitable for localization by ASLAD.

increasing frequency, the chirp provides a unique signal suitable for localization by ASLAD. The higher frequency components, after filtering, result in a mean value of noise level. The lower frequency components, after filtering, produce a unique signal. The chirp signal, target signal, and data acquisition interval are all periodic. However all three have different periods, and as a result, the target and noise signals change with respect to one another in a way that does not occur with the white noise source. For any given data acquisition interval, the pulse of the target signal and the low frequency components of the chirp signal may or may not overlap, modifying the target signal in a different way for each data acquisition interval. Or, the target signal may not be present at all in the data acquisition interval. By itself, the chirp, presented equidistant from all three omnidirectional microphones, provides a signal that could be successfully localized. However, it would be rejected in the desired position function, as all three omnidirectional microphone pairs would qualify as the primary pairing.

The second series of measurements were done using a chirp signal as the noise. The range of frequencies for the chirp was 250 Hz to 3.1 kHz, and the length of the sweep cycle was approximately 0.1 seconds. This noise signal was chosen because of its unique properties as described above. It was predicted that a complicated noise signal such as this would detrimentally modify the target signal, and thus a higher signal-to-noise ratio would be required for ASLAD to function. For this experiment, the setup and test source were the same as for the white noise experiment, with the exception that the target signal was set to an SPL of 77.8 ± 0.5 dB. The maximum available SPL for the noise signal was 70.4 ± 0.5 dB. The background SPL, with no noise or target signal present, was 50.7 ± 0.5 dB. As in the white noise test, the SPL of the noise signal was reduced in 3 to 4 dB steps, and the performance of ASLAD was recorded at each step.

Below a signal-to-noise ratio of 18 dB SPL, ASLAD could not localize the target, and

the directional microphone was in constant, random motion. At a signal-to-noise ration of 18 ± 1 dB SPL, the directional microphone was repeatedly moving to the target angle, stopping briefly, then moving about randomly. At a signal-to-noise ratio of 22 ± 1 dB SPL, the directional microphone was able to lock onto the target angle, but was slow to find the correct location. At a signal-to-noise level of 24 ± 1 dB SPL, ASLAD was working as well as it did with no noise present. As predicted, with this very complicated noise signal, a signal-to-noise ratio of greater than 20 dB SPL was required in order for the sound localizing device to work properly.

4.2.4 Time Required to Locate and Move to a Target

The time required to locate and move to a new target is important. This tells us what length of delay exists between the time that a new target is introduced and the time that it would be picked up by the directional microphone. It also gives some indication about how many cycles of thesound localizing algorithm are required, on the average, to identify a new target. For all tests, the new target was presented at a position 90° offset from the existing position of the directional microphone. A rotation of the directional microphone through this angle takes approximately 0.4 seconds. Three different tests were completed. For the first two tests, introduction of the target signal was preceded by a period of silence (approximately 15 seconds of silence). For the third test, two target speakers were set up at a separation of 90°, and presentation of a signal was switched from one speaker to the other at the same instant. For the first two tests, one target signal was live voice, and the other was the AM sine signal (as previously defined). For the third test, two target signals generated by function generators were used.

Table 4.4 lists the results of the timing experiments. From this table it can be seen that the voice target resulted in the lowest relocation times (approximately 1.5 seconds). Subtracting

0.4 seconds to account for the travel of the directional microphone, this leaves 1.1 seconds required for calculation of the target angle. From this it can be concluded that, on average, 6 cycles of thesound localizing algorithm (where one cycle is 177 milliseconds long: 100 millisecond of data collection and 77 milliseconds for the execution of thesound localizing algorithm) are required to locate a voice source. The test to measure time required to locate a source from a silent environment was repeated with the AM sine signal. The results of this test indicated that it took, on average, 11 cycles of thesound localizing algorithm to identify the sound source. In comparison, the results of the third test indicated that, when switching from one target to another, an average of 20 cycles of thesound localizing algorithm were necessary.

The longer switching time for the third test is the effect of the delay incorporated into thesound localizing algorithm. As described in section 3.5.2, this delay was introduced to reduce number of instances where the successful localization of a target was interrupted by the localization of a transient target. While the weighted averaging of the solution array did succeed in reducing the effect of transient noises on directional microphone movement, it also introduced a delay in the acquisition of desired target signals. The implications of the delay in the localization of new targets on the quality of the signal collected from the directional microphone will be discussed in section 5.1.

These results are consistent with expectations. The number of cycles of the sound localizing algorithm required to "lock" on a target could possibly be reduced by increasing the time period for the collection of each set of data from the omnidirectional microphones. This would increase the chances of collecting a sound sample in which a clear peak would be present (see section 4.1.1). However, if the same sampling rate is to be maintained to preserve the localization resolution, the data collection period, and hence the length of the data arrays, must be increased. Any increase in the length of the data arrays will also increase the processing time.

To use the FFT, the array must be of length 2^n ($n \in I$). The data arrays would have to be increased to 4096 samples. The processing time of 4096 point FFTs and inverse FFTs is not two, but three to four times longer than for 2048 points, due to the limited high speed RAM on the DSP chip. Unless the increased data array length resulted in the average number of cycles required for localization of a target to be lowered to two, the average time required to locate and move to a new target would increase instead of decrease.

4.2.5 Behavior when Presented with Two Targets Simultaneously

To observe the behaviour of ASLAD when presented with two targets simultaneously, two different tests were performed. For one test, two different targets were presented at 90° separation, to ASLAD. One target was a continuous sine signal of frequency 1 kHz and 77 dB SPL. The other target was the AM sine signal. The continuous sine target was localized if the AM sine target was at a lower SPL and also when the AM sine signal was up to 5 dB higher than the continuous sine target. Timing tests, performed in the same manner as described in the previous section, showed that the time required to locate the continuous sine target was lower than for the AM sine target. In other words, not every data acquisition interval resulted in the collection of data for which the pulsed target could be localized. Due to the weighted averaging of past solutions, the more easily localized target was being chosen as the target at which the directional microphone should be aimed.

For the second test, continuous sine targets of the same frequency and SPL were simultaneously presented at two different speakers separated by an angle of 90°. The result was that ASLAD was very unstable. The directional microphone searched constantly between the angles of the two targets and was not able to settle at any one angle. The two continuous sine targets were generated using two different function generators. If the frequency from one was

not exactly the same as the other, as would most likely be the case, the envelope of the two signals would always be changing relative to one another. If the two signals are constantly changing relative to one another due to a slight difference in the period of each, then it stands to reason that the solution to the localization algorithm will change with every cycle.

4.3 Qualitative Evaluation of the Sound Localizing Device in an Uncontrolled Acoustic Environment

In its current form, ASLAD can successfully locate target sounds and steer the directional microphone to point at the identified target. In an environment where acoustic conditions could be described, at best, as fair, ASLAD responds convincingly to sharp noises such as finger snapping, and also to pure tones of frequencies within the bandwidth of the digital low-pass filter used to obtain the envelope of the collected sound samples. ASLAD also responds well to a single person moving about the device and talking, as well as to two (or more) people having a conversation. As described above, there is a delay built into the system to reduce the effect of transient noise, so for the conversation to be tracked well by the system the "volley" of conversation must be slow. That is, best results are achieved if each person talks for a significant period of time (i.e. several sentences) without interruption.

ASLAD does, however, occasionally move the directional microphone to the wrong position, move it when no new target has been presented or when an existing target is still active. It is interesting to note that when this happens, it often moves to a position 180° opposite to the position of the target. The 180° reversals occurred more frequently in the uncontrolled acoustic environment than it did in the anechoic chamber. This provides further evidence that a portion of the localizing errors are due to echoes. Dealing with echoes is an important issue that must be considered in future versions of ASLAD if it is to be robust in less than ideal acoustic

conditions. In its present form, ASLAD is also easily confused by the presence of multiple targets. Again, this is something that would have to be taken into account in the future development of ASLAD prototypes.

CHAPTER 5: Discussion and Conclusions

5.1 Discussion and Recommendations for Future Work

For an assistive listening device to be useful it must, in addition to enhancing speech intelligibility for the user, be small enough and sturdy enough to be portable, and should not disturb others. In order to make ASLAD portable, it would be necessary to move a power supply (either batteries or transformer) into the case, and design a custom processing card with the required DSP chip, memory, program, A/D and D/A converters on it. The directional microphone would also have to be made easily removable. The directional microphone is currently powered by its own battery. In a stand alone device, the power for this microphone could be tapped from the main power supply. The size of ASLAD would be comparable to that of a laptop computer. The platform assembly rotates silently, and has been made as small as possible to minimize distractions it may cause. Nevertheless, the motion of the directional microphone may still be distracting to some people.

From an aesthetic point of view, it would be preferable if the directional microphone could be concealed. This would minimize the distraction created by the device. People with disabilities are often self-conscious, and they do not like to draw attention to themselves or make other people feel uncomfortable about interacting with them. Unfortunately it would be very difficult to conceal the directional microphone in its present form. Any covering used to conceal the directional microphone would shadow one or more of the omnidirectional microphones, causing problems for the sound localizing algorithm. In addition, the material used to conceal the directional microphone could affect the quality of the collected signal. If, however, a suitable material could be found that would not alter the signal collected by the directional microphone, it would be possible to conceal it. For example, by separating the omnidirectional microphone

array and the DMPS into two separate units, the DM could be concealed in some sort of case without shadowing the omnidirectional microphones. This would require some recalibration of the localization algorithm to account for the offset between the omnidirectional microphone array and the DMPS. Concealing the microphone without affecting the sound quality may not turn out to be possible, but it is worth exploring this option.

The spectral characteristics of the signals acquired by the omnidirectional microphones are not analyzed in the sound localizing algorithm. However, these characteristics are important to keep in mind for future versions of ASLAD. The spectral characteristics may be useful for filtering out sounds on the basis of whether or not they are speech. For instance, in the sound localizing algorithm, one of the first steps is to low-pass filter the data to obtain the envelope of each signal acquired by the omnidirectional microphones. As a Fourier Transform method is used to perform the low pass filtering digitally, a routine could easily be written to make a copy of the transformed data and analyze the spectral content of the incoming signal. As an example, energy density in the frequency range between 1 and 2 kHz may be a good indication of whether the signal is speech or non-speech. Figures 3(a) and 3(b) are examples of the spectra of speech and non-speech sources. A number of speech and non-speech samples were tested using a commercial sound recording and analysis software package. There tended to be very little energy in the 1 to 2 kHz range of the spectrum for the speech samples as compared to the non-speech samples. Alternatively, a subroutine could be written to identify harmonics in the collected signals. As shown in Figure 3(a), there are a distinctive set of harmonics present in the spectrum for the speech sample.

Upon implementation of the sound localizing algorithm, it was determined that it would be necessary to incorporate a delay into the routine to help remove the effect of transient noise and echoes. This is done by considering a weighted average of past results of the sound

localizing algorithm in addition to each new result. Without this feature, the directional microphone moved around with each cycle of the sound localizing algorithm. The delay does succeed in stabilizing the motion of the directional microphone, but also slows the response time of the directional microphone to acquisition of desired targets. This means that if two people are having a conversation, there will be a delay before a change in the speaker is recognized. By setting the weighting factor for averaging of the past results, a delay of any number of cycles of the algorithm can be set. By experiment it was found that a minimum delay of 3 cycles of data acquisition and execution of the sound localizing algorithm was necessary to observe a reduction in directional microphone motion. A delay of 5 cycles was found to be suitable balance for achieving stable performance without introducing an excessive delay. A 5 cycle delay was used for the tests performed in the anechoic chamber.

In its current form, the only sound signal sent to the user of this hearing assistive device is that from the directional microphone. The directional microphone used in ASLAD is designed specifically to be connected to a particular type of hearing aid. When connected, the signal from the microphone in the hearing aid is automatically attenuated by 6 dB, and can be switched to 18 dB attenuation. Because the time required for the platform to rotate can be as much as 0.9 seconds, and because it was necessary to incorporate a delay into the sound localizing algorithm to make the DMPS more stable and less prone to transient noises, there will be times when the directional microphone is not pointing in the correct direction. Indeed, from the results of the timing tests performed in the anechoic chamber, it was clear that there would be cases where three to four seconds could elapse before the directional microphone settled at the new location. For the existing setup, when the directional microphone not pointing at the desired target, the user may still be able to hear the desired source via the hearing aid microphones. However, the signal would be attenuated. To create a system independent of the hearing aid, or to bypass the

hearing aid microphone altogether, it would be necessary to design a method of switching between the signal from the directional microphone and the signal from one of the omnidirectional microphones. The intensity of this signal would have to be adjusted to match that from the directional microphone so that switching would not result in disturbing level changes for the listener. An algorithm would have to be developed to determine which output signal should be directed to the hearing aid. For instance, one might decide that if the desired position of the platform changes by more than a threshold value, then the signal from the omnidirectional microphone should be sent to the hearing aid, as the new target must be out of the range of the directional microphone. Also, a confidence level could be established, and if the sound localizing algorithm does not meet the confidence level then the signal from the omnidirectional microphone should be sent to the hearing aid. This confidence level could use factors such as the number of consecutive cycles of the sound localizing algorithm giving the same result, or the number of cycles of the sound localizing algorithm giving conflicting results in a certain time period, as an indication of the certainty that the directional microphone is pointing in the correct direction.

Switching between the signal from the directional microphone and the omnidirectional microphone as the feed for the user is an interesting problem, and one that must be dealt with in future versions of ASLAD. This is something that could be done entirely in software, as a part of the control algorithm. The goal of an algorithm designed to do this would be to maximize the amount of time that the directional microphone could be used, while ensuring that no information was lost when the position of the target was either changing or uncertain.

The most likely situation where ASLAD would be useful is one where there is a small group of people sitting around a table, with only one person talking at any one time. In such a situation where the targets were stationary, it may be advantageous to give the control algorithm

some sort of memory. For instance, the positions that the directional microphone was most often sent to could be kept track of and trajectories for each of the possible moves from one common position to another could be calculated. This could reduce the time delay in switching from one target to another, and could also save processing time for the sound localizing algorithm. If ASLAD was used to monitor a meeting between two people, it would quickly become evident that the directional microphone needs to alternate between two stationary targets. A scaled back version of the sound localizing algorithm could then be employed part of the time to find only general direction and send the directional microphone to the previously identified target in that general direction. The full sound localizing algorithm would act as a periodic monitor to accommodate changes in the situation such as one of the people moving, or an additional person entering the conversation.

The DMPS has only one degree of freedom: rotation about a vertical axis. For the characteristics of the directional microphone being used and the intended nature of use this is adequate, as most people sitting or standing around the microphone would fall within its cone of acceptance. However if the directional microphone's acceptance angle was small enough, or if additional flexibility was required, a second degree of freedom would be required. This would be relatively easy to add. An additional omnidirectional microphone would have to be positioned above the existing array to be able to perform the sound localizing algorithm for the vertical plane. The signal processing would be an extension of that already in place, and the platform would have to be modified to incorporate an additional actuator. One simple way of achieving motion in the vertical plane would be to pivot the directional microphone at its balance point, and attach a linear actuator at its back end. However, the design of the platform to include an additional actuator could prove to be a challenge if the directional microphone had to be easily detachable for storage and transportation. To add this second degree of freedom, one additional

A/D converter would be required for the extra omnidirectional microphone, with possibly a second A/D converter if position feedback was required for the linear actuator. One additional D/A converter would be required for the controlling signal for the linear actuator.

The control algorithm implemented in this prototype is simple, but it works well. Due to the computationally intensive sound localizing algorithm, it was not possible to sample the environment at the desired rate and collect the position feedback signal continuously without significant increases in the processing time of the sound localizing algorithm. This would have been required in order to add a derivative term to the control routine. The derivative term would be useful in the reduction of the overall travel time. By knowing the velocity of the platform, the overshoot could be anticipated and minimized by initiating the actions necessary to bring the platform to a halt sooner. If it is eventually possible to have directional microphone of the size used in ASLAD with a "pencil beam" cone of acceptance, the simple control schemes used in ASLAD would not suffice. In this event, it would be desirable to obtain position data at fixed intervals so that the directional microphone could be made to settle at a given position with a minimum of overshoot and oscillation about that point.

The improved behaviour of ASLAD with respect to random motion of the directional microphone and incorrect localizations in the anechoic environment is a strong indicator that echoes negatively affect its performance. Also, in the case where two target signals, equally likely to be localized successfully, were simultaneously presented to ASLAD from different angles, ASLAD was unable to find either target, and moved around almost constantly during the experiment. For the device to be useful in everyday life, the problem of dealing with echoes and multiple sources in the sound localization process would have to be addressed.

Perhaps the most interesting, and certainly the most challenging work to be considered in future versions of this device is the refinement of the sound localizing algorithm. The ability

to deal with multiple sources and echoes, and to discriminate between speech and non-speech sounds would increase the number of situations where this device would be useful. The manipulation of speech signals can be very processor-intensive. The need to work in real time will always be a limiting factor in the complexity of the sound localizing algorithm used.

5.2 Summary and Conclusions

The objective of this project was to design and develop a prototype of a new type of hearing assistive device. The concept called for the use of data collected from a set of omnidirectional microphones to determine the location of sound sources, and for an automated directional microphone that moved to point at an identified source as the means of acquiring the desired signal with a high signal to noise ratio. The first step of this project was to develop a conceptual model for the device, identify the key stages in the model, and identify some of the components in each key stage. Figure 1 is a flowchart which shows the four key stages of the conceptual model, as well as the probable flow of information through the system. Next, it was necessary to derive a set of specifications for each of the different stages of the device, as well as for the interface between stages. These specifications were based on the performance expected from such an assistive listening device, and on the desire to build a system to which new sound localizing algorithms could easily be adapted in the future. The specifications were developed for an array of two omnidirectional microphones. The design was based on these specifications.

In addition to building ASLAD to perform to the specifications derived for the prototype, several other factors had to be considered in the physical design of the device. As outlined in the introductory chapter of this thesis, a personal hearing assistive device must be portable, robust, and must operate silently. Furthermore, the localization resolution of the sound localizing algorithm is dependent on the omnidirectional microphone separation and the sampling rate.

Finally, the processing time required to execute the sound localizing algorithm and the DMPS control algorithm had to be minimized wherever possible, in order to develop a system that would operate in real time. The process of balancing each of these factors in the design process to meet as many of the preliminary specifications as possible is described in detail in chapter three.

During the course of construction of the device, some changes were made to the design that resulted in changes to the specifications. Most notable was the decision to use three omnidirectional microphones instead of two for the purposes of sound localization. The result was that there were three combinations of two omnidirectional microphones to which the sound localizing algorithm could be applied, giving the ability to localize sound over a full 360° range. The change to a three omnidirectional microphone system allowed the construction of a device that was still portable, but more robust, with fewer moving parts, and with a directional microphone platform that was easier to control. Although processing time for the sound localizing algorithm was increased by 75% with the addition of the third omnidirectional microphone, the benefits gained in terms of localizing ability, robustness, and ease of directional microphone control justified the increased processing time. Other than specifications affected by this change, and the decision to use a lower sampling rate than initially chosen to help keep processing time down, all original specifications were met in the prototype.

The finished device consists of an array of three omnidirectional microphones arranged in an equilateral triangle, at the geometric centre of which is a directional microphone mounted on a belt-driven rotating platform. Sound samples are collected from the omnidirectional microphones and stored on a digital signal processing board. A sound localizing algorithm is applied to the data. The algorithm calculates the position to which the directional microphone is to be moved. This information is passed onto the DMPS, which evaluates the existing and

desired positions of the directional microphone, and sends an appropriate signal to the motor that drives the rotating platform on which the directional microphone is mounted.

The directional microphone has a range of 240°, through which it can rotate in under one second, and settle on the chosen location with an error of less than 5°. It is able to locate and focus on sound sources in a quiet environment where only one source is presented at a time. The device operates silently, and is compact enough to be called portable.

Tests performed in an anechoic environment showed that the device could localize sound sources with an average error of approximately 5°. The behaviour of the device with respect to random motion of the directional microphone and incorrect localizations in the anechoic environment was much improved, which is a strong indicator that echoes negatively affect the performance of the device. The device could function normally in the presence of white noise in the anechoic chamber with a 15 dB signal to noise ratio. Time delays between the onset of a new sound source and the settling of the directional microphone at the position of the new source were measured to be as high as 4 seconds in some conditions, suggesting that there is a need to develop a method of switching between the directional and omnidirectional microphones for the collection of the signal to prevent the loss of information during the transition period between the loss of one target and the identification of new one.

The construction and testing of this device has been very helpful in the identification of problems that were not obvious on paper. Only then can one see the behaviour of the system as a whole. The most notable example is the need for a delay term in the sound localizing algorithm. The need to slow the process of identifying new targets in order to minimize the effect of transient noises and echoes on the control of the DMPS was only evident once there was a moving object to be controlled.

For an assistive listening device to be useful, it must make an improvement over the

ability of the user to function without it. There are two major issues that need to be addressed in future versions of this device to improve its performance, and usefulness, as an assistive listening device. They are: the ability to localize and select one target in the presence of multiple targets; and the ability to switch between the signal from the directional microphone and the omnidirectional microphones when a single target cannot be identified, or during a position change of the directional microphone. In its present form, ASLAD is easily confused by multiple targets, and hence would only be useful in controlled environments with orderly conversation. When the directional microphone is correctly aimed at a target source, the characteristics of the directional microphone ensure that there would be some improvement in the signal to noise ratio of the acquired signal. However, when the directional microphone is not correctly aimed at the target, the user would receive very little of the desired sound signal, resulting in a reduction of the signal to noise ratio as compared to that acquired without the assistive listening device.

The software for ASLAD is written as a series of functions that are called repeatedly. This makes it possible to change any one algorithm without affecting the rest of the program. It also makes it possible to use existing functions as building blocks for the addition of new algorithms. It is hoped that in the future, this device can be used as a testbed for new sound localizing algorithms and noise reduction techniques.

REFERENCES

Bhadkamkar, N., Fowler, B., A Sound Localization System Based On Biological Analogy. IEEE International Conference on Neural Networks. San Francisco, March 1993. Vol. 3, pp. 1902-7.

Butler, R.A., Monaural and binaural localization of noise bursts vertically in the median sagital plane. Journal of Auditory Research, vol. 3 pp. 230-5. 1969.

Colburn, H.S., Computational Models of Binaural Processing. (Chapter for Auditory Computation, Hawkins, H. and McMullen, T. eds.) in press.

Davis, A.C., The Prevalence of Hearing Impairment and Reported Hearing Disability Among Adults in Great Britain. International Journal of Epidemiology, 18(4) pp. 911-7, 1989.

Duquesnoy, A.J., Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. Journal of the Acoustical Society of America 74(3) pp. 739-43, September 1983.

Engebretson, A.M., Benefits of Digital Hearing Aids. IEEE Engineering in Medicine and Biology, 13(2) pp. 238-48, April/May 1994.

Feder, M., Oppenheim, A.V., Weinstein, E., Maximum Likelihood Noise Cancellation Using the EM Algorithm. IEEE Transactions on Acoustics, Speech, and Signal Processing 37(2) pp. 204-16, February 1989.

Flanagan, J.L., Johnston, J.D., Zahn, R., Elko, G.W., Computer steered microphone arrays for sound transductions in large rooms. Journal of the Acoustical Society of America 78(5) pp. 1508-18, November 1985.

Franklin, G.f., Powell, J.D., Emami-Naeini, A., Feedback Control of Dynamic Systems. Addison-Wesley Publishing Company, Inc., Menlo Park, California. 1986.

Fry, D.B., The Physics of Speech. Cambridge University Press, Cambridge. 1979.

Greenberg, J.E., Zurek, P.M., Evaluation of an adaptive beamforming method for hearing aids. Journal of the Acoustical Society of America 91(3) pp. 1662-76, March 1992.

Haykin, S., An Introduction to Analog and Digital Communications. John Wiley & Sons, Inc., New York. 1989.

Hoffman, M.W., Trine, T.D., Buckley, K.M., Van Tasell, D.J., Robust Adaptive Microphone Array Processing for Hearing Aids: Realistic Speech Enhancement. Journal of the Acoustical Society of America, 96(2) pp. 759-70, August 1994.

Jeffress, L.A., A Place Theory of Sound Localization. J Comp Physiol Psychol, vol. 41 pp. 35-9. 1948.

Jung, W.C., I.C. Op Amp Cookbook, 1st edition. Howard W. Sams & Co. Inc., Indianapolis, 1974.

Kates, J.M., Superdirective Arrays for Hearing Aids. Journal of the Acoustical Society of America, 94(4) pp. 1930-3, October 1993.

Levitt, H., Weiss, M., Neuman, A., Noise Reduction for Hearing Aids. Proceedings of the 15th Annual Northeast Bioengineering Conference(IEEE), Boston, March 1989. p. 25.

Link, M.J., Hoffman, M.W., Buckley, K.M., Hardware Implementation of Robust Constrained Hearing Aid Arrays. Conference Record of the 26th Asilomar Conference on Signals, Systems & Computers (IEEE), Pacific Grove, October 1992. Vol. 1, pp. 540-4.

Middlebrooks, J.C., Green, D.M., Directional Dependence of interaural envelope delays. Journal of the Acoustical Society of America 87(5) pp. 2149-62, May 1990.

Mills, A.W., On the minimum audible angle. Journal of the Acoustical Society of America 30: 237-246, April 1958.

Ogata, K., Modern Control Engineering, second edition. Prentice Hall, Englewood Cliffs, New Jersey. 1990.

Oppenheim, A.V., Schafer, R.W., Discrete-Time Signal Processing. Prentice Hall, Englewood Cliffs, New Jersey. 1989.

Palmieri, F., Shah, A., Moiseff, A., Neural Coding of Interaural Time Difference. IJCNN International Joint Conference on Neural Networks (IEEE). Baltimore, June 1992. Vol. 4, pp. 271-6.

Papamichalis, P. (ed.), DSP Applications with the TMS320 Family, volume 3. Texas Instruments, Inc. 1990.

Plomp, R., Noise, Amplification, and Compression: Considerations of Three Main Issues in Hearing Aid Design. Ear & Hearing, 15(1) pp. 2-12, February 1994.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., Numerical Recipes. Cambridge University Press, Cambridge, Mass. 1986.

Proakis, J.G., Manolakis, D., Introduction to Digital Signal Processing, second edition. Macmillan Publishing Company. 1992.

Sammeth, C.A., Ochs, M.T., A Review of Current "Noise Reduction" Hearing Aids: Rationale, Assumptions, and Efficacy. Ear and Hearing 12(6) Supplement, pp. 116S-24S. 1991.

Schwander, T. Levitt, H., Effect of two-microphone noise reduction on speech recognition by normal-hearing listeners. Journal of Rehabilitiation Research and Development 24(4) pp. 87-92, Fall 1987.

Searle, C.L., Braida, L.D., Davis, M.F., Colburn, H.S., Model for auditory localization. Journal of the Acoustical Society of America 60(5) pp. 1164-75. November 1976.

Shein, J. (ed.), Canadians with Impaired Hearing, Publication Division, Statistics Canada, 1992.

Sorensen, H.V., Jones, D.L., Heideman, M.T., Burrus, C.S., Real-valued FFT algorithms. IEEE Transactions on Acoustics, Speech, and Signal Processing 35(6) pp. 849-63, June 1987.

Soede, W., Berkhout, A.J., Bilsen, F.A., Development of a Directional Hearing Instrument Based on Array Technology. Journal of the Acoustical Society of America, 94(2), Pt. 1, pp. 785-98. August 1993.

Soede, W., Bilsen, F.A., Berkhout, A.J., Assessment of a Directional Microphone Array for Hearing Impaired Listeners. Journal of the Acoustical Society of America, 94(2), Pt. 1, pp. 799-808. August 1993.

Sugie, N., Huang, J., Ohnishi, N., Localizing Sound Source by Incorporating Biological Auditory Mechanism. IEEE International Conference on Neural Networks. San Diego, July 1988. Vol. 2, pp. 243-50.

Van De Graaff, K.M., Human Anatomy, second edition. Wm. C. Brown Publishers, Dubuque, Iowa. 1988.

Weiss, M., Use of an adaptive noise canceler as an input preprocessor for a hearing aid. Journal of Rehabilitation Research and Development 24(4) pp. 93-102, Fall 1987.

Yuhas, B.P., Automated Sound Localization Through Adaptation. IJCNN International Conference on Neural Networks (IEEE). Baltimore, June 1992. Vol. 2, pp. 907-12.

Zakarauskas, P., Patent Application, 1993.