THE UNIVERSITY OF CALGARY

Localization of Sound Using Headphones

by

Edward C. Beingessner

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

CALGARY, ALBERTA

JANUARY, 1994

ISBN 0-315-93860-9

Name Edward Beingessner

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Electronics and Electrical Engineering
SUBJECT TERM

0544
SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology ... 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology .. 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences

Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics ... 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

Nom _____

Dissertation Abstracts International est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.

_____ ☐☐☐☐ U·M·I

SUJET CODE DE SUJET

Catégories par sujets

HUMANITÉS ET SCIENCES SOCIALES

COMMUNICATIONS ET LES ARTS
Architecture 0729
Beaux-arts 0357
Bibliothéconomie 0399
Cinéma 0900
Communication verbale 0459
Communications 0708
Danse 0378
Histoire de l'art 0377
Journalisme 0391
Musique 0413
Sciences de l'information 0723
Théâtre 0465

ÉDUCATION
Généralités 515
Administration 0514
Art 0273
Collèges communautaires 0275
Commerce 0688
Économie domestique 0278
Éducation permanente 0516
Éducation préscolaire 0518
Éducation sanitaire 0680
Enseignement agricole 0517
Enseignement bilingue et multiculturel 0282
Enseignement industriel 0521
Enseignement primaire. 0524
Enseignement professionnel 0747
Enseignement religieux 0527
Enseignement secondaire 0533
Enseignement spécial 0529
Enseignement supérieur 0745
Évaluation 0288
Finances 0277
Formation des enseignants 0530
Histoire de l'éducation 0520
Langues et littérature 0279
Lecture 0535
Mathématiques 0280
Musique 0522
Orientation et consultation 0519
Philosophie de l'éducation 0998
Physique 0523
Programmes d'études et enseignement 0727
Psychologie 0525
Sciences 0714
Sciences sociales 0534
Sociologie de l'éducation 0340
Technologie 0710

LANGUE, LITTÉRATURE ET LINGUISTIQUE
Langues
- Généralités 0679
- Anciennes 0289
- Linguistique 0290
- Modernes 0291

Littérature
- Généralités 0401
- Anciennes 0294
- Comparée 0295
- Médiévale 0297
- Moderne 0298
- Africaine 0316
- Américaine 0591
- Anglaise 0593
- Asiatique 0305
- Canadienne (Anglaise) 0352
- Canadienne (Française) 0355
- Germanique 0311
- Latino-américaine 0312
- Moyen-orientale 0315
- Romane 0313
- Slave et est-européenne 0314

PHILOSOPHIE, RELIGION ET THÉOLOGIE
Philosophie 0422
Religion
- Généralités 0318
- Clergé 0319
- Études bibliques 0321
- Histoire des religions 0320
- Philosophie de la religion 0322

Théologie 0469

SCIENCES SOCIALES
Anthropologie
- Archéologie 0324
- Culturelle 0326
- Physique 0327

Droit 0398
Économie
- Généralités 0501
- Commerce-Affaires 0505
- Économie agricole 0503
- Économie du travail 0510
- Finances 0508
- Histoire 0509
- Théorie 0511

Études américaines 0323
Études canadiennes 0385
Études féministes 0453
Folklore 0358
Géographie 0366
Gérontologie 0351
Gestion des affaires
- Généralités 0310
- Administration 0454
- Banques 0770
- Comptabilité 0272
- Marketing 0338

Histoire
- Histoire générale 0578
- Ancienne 0579
- Médiévale 0581
- Moderne 0582
- Histoire des noirs 0328
- Africaine 0331
- Canadienne 0334
- États-Unis 0337
- Européenne 0335
- Moyen-orientale 0333
- Latino-américaine 0336
- Asie, Australie et Océanie 0332

Histoire des sciences 0585
Loisirs 0814
Planification urbaine et régionale 0999
Science politique
- Généralités 0615
- Administration publique 0617
- Droit et relations internationales 0616

Sociologie
- Généralités 0626
- Aide et bien-âtre social 0630
- Criminologie et établissements pénitentiaires 0627
- Démographie 0938
- Études de l' individu et de la famille 0628
- Études des relations interethniques et des relations raciales 0631
- Structure et développement social 0700
- Théorie et méthodes. 0344
- Travail et relations industrielles 0629

Transports 0709
Travail social 0452

SCIENCES ET INGÉNIERIE

SCIENCES BIOLOGIQUES
Agriculture
- Généralités 0473
- Agronomie. 0285
- Alimentation et technologie alimentaire 0359
- Culture 0479
- Élevage et alimentation 0475
- Exploitation des pâturages 0777
- Pathologie animale 0476
- Pathologie végétale 0480
- Physiologie végétale 0817
- Sylviculture et faune 0478
- Technologie du bois 0746

Biologie
- Généralités 0306
- Anatomie 0287
- Biologie (Statistiques) 0308
- Biologie moléculaire 0307
- Botanique 0309
- Cellule 0379
- Écologie 0329
- Entomologie 0353
- Génétique 0369
- Limnologie 0793
- Microbiologie 0410
- Neurologie 0317
- Océanographie 0416
- Physiologie 0433
- Radiation 0821
- Science vétérinaire 0778
- Zoologie 0472

Biophysique
- Généralités 0786
- Medicale 0760

SCIENCES DE LA TERRE
Biogéochimie 0425
Géochimie 0996
Géodésie 0370
Géographie physique 0368
Géologie 0372
Géophysique 0373
Hydrologie 0388
Minéralogie 0411
Océanographie physique 0415
Paléobotanique 0345
Paléoécologie 0426
Paléontologie 0418
Paléozoologie 0985
Palynologie 0427

SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT
Économie domestique 0386
Sciences de l'environnement 0768
Sciences de la santé
- Généralités 0566
- Administration des hôpitaux 0769
- Alimentation et nutrition 0570
- Audiologie 0300
- Chimiothérapie 0992
- Dentisterie 0567
- Développement humain 0758
- Enseignement 0350
- Immunologie 0982
- Loisirs 0575
- Médecine du travail et thérapie 0354
- Médecine et chirurgie 0564
- Obstétrique et gynécologie 0380
- Ophtalmologie 0381
- Orthophonie 0460
- Pathologie 0571
- Pharmacie 0572
- Pharmacologie 0419
- Physiothérapie 0382
- Radiologie 0574
- Santé mentale 0347
- Santé publique 0573
- Soins infirmiers 0569
- Toxicologie 0383

SCIENCES PHYSIQUES
Sciences Pures
Chimie
- Généralités 0485
- Biochimie 487
- Chimie agricole 0749
- Chimie analytique 0486
- Chimie minérale 0488
- Chimie nucléaire 0738
- Chimie organique 0490
- Chimie pharmaceutique 0491
- Physique 0494
- PolymÇres 0495
- Radiation 0754

Mathématiques 0405
Physique
- Généralités 0605
- Acoustique 0986
- Astronomie et astrophysique 0606
- Électronique et électricité 0607
- Fluides et plasma 0759
- Météorologie 0608
- Optique 0752
- Particules (Physique nucléaire) 0798
- Physique atomique 0748
- Physique de l'état solide 0611
- Physique moléculaire 0609
- Physique nucléaire 0610
- Radiation 0756

Statistiques 0463

Sciences Appliqués Et Technologie
Informatique 0984
Ingénierie
- Généralités 0537
- Agricole 0539
- Automobile 0540
- Biomédicale 0541
- Chaleur et ther modynamique 0348
- Conditionnement (Emballage) 0549
- Génie aérospatial 0538
- Génie chimique 0542
- Génie civil 0543
- Génie électronique et électrique 0544
- Génie industriel 0546
- Génie mécanique 0548
- Génie nucléaire 0552
- Ingénierie des systâmes 0790
- Mécanique navale 0547
- Métallurgie 0743
- Science des matériaux 0794
- Technique du pétrole 0765
- Technique minière 0551
- Techniques sanitaires et municipales 0554
- Technologie hydraulique 0545

Mécanique appliquée 0346
Géotechnologie 0428
Matières plastiques (Technologie) 0795
Recherche opérationnelle 0796
Textiles et tissus (Technologie) 0794

PSYCHOLOGIE
Généralités 0621
Personnalité 0625
Psychobiologie 0349
Psychologie clinique 0622
Psychologie du comportement 0384
Psychologie du développement 0620
Psychologie expérimentale 0623
Psychologie industrielle 0624
Psychologie physiologique 0989
Psychologie sociale 0451
Psychométrie 0632

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "Localization of Sound Using Headphones" submitted by Edward C. Beingessner in partial fulfillment of the requirements for the degree of Master of Science.

Supervisor, Dr. L. E. Turner
Electrical and Computer Engineering

Dr. M. Fattouche
Electrical and Computer Engineering

Dr. J. Kendall
Computer Science

Dr. E. Nowicki
Electrical and Computer Engineering

10 January 1994
(Date)

Abstract

When listening to prerecorded sound on headphones, the listener perceives that the sounds are located inside the head. This thesis investigates ways of altering the perceived location of sounds during headphone listening. Sound locations are moved outside the head and into the horizontal plane.

Sound recordings are conventionally listened to in a room using two speakers placed in front of the listener. A means of simulating a conventional listening environment with headphones, using head related transfer functions [Bla83] and reverberation [Moo79], is described and realized in hardware.

This method proved inadequate for the simulation of a multi-speaker system. In order to stimulate accurate localization in the horizontal plane, a listener must adapt to the simulated environment. A means of using revolving sounds to allow a listener to adapt to a multi-channel environment is presented.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Music is one of the greatest forms of entertainment. People enjoy relaxing with their favorite recordings. There are stereos in almost every home, and some people take small personal stereos wherever they go. Many methods of storing recorded music exist (records, tapes, compact discs, etc.), but there is only one method of producing sound: speakers. Speakers change recordings back into sound.

How the speakers are arranged is important; it effects the way the music is heard. Two distinct speaker arrangements are used: loudspeaker arrangements and headphones arrangements. A loudspeaker arrangement plays sounds aloud for all to hear. It consists of two (or more) speakers placed at some distance from the listener. A headphone arrangement, the subject of the thesis, uses two speakers placed on either side of the head, one covering each ear.

Headphones are popular because they are quieter than loudspeakers. Loudspeakers contribute to sound pollution by projecting sound into the general environment, whereas headphones project sound discretely into the ears of the listener. There are, however, problems with the way headphones introduce sound to the listener.

Headphones sound unnatural. For example, headphones can produce a sound heard in only one ear. In nature, sound from a source radiates in all directions and will propagate to both of the listener's ears; thus, nothing in nature has prepared the human auditory system to hear a sound in only one ear. Thus, headphones sound confusing.

Headphone presentation also has its benefits. Headphones completely control what the listener hears. The sound presented to each ear is controlled independently and the surrounding environment does not effect the presentation. In contrast, the

sound from a loudspeaker is heard in both ears, and it is altered by the environment. Surfaces in the environment surrounding the listener, like walls and ceilings, will reflect the sound from a loudspeaker. These reflections will alter what is heard by the listener. Headphones avoid these reflections by playing directly into the ear.

Overcoming the problems and exploiting the advantages of headphone listening is the goal of this thesis. Before proceeding further, a common vocabulary for dealing with sound, systems and hearing must be defined. Sound is discussed first.

1.1 Sounds, Digital Signals, and Frequency Components

Sound is a pattern of pressure differences transmitted through a medium as a wave. The pressure differences stimulate a listener's ear and provoke hearing.

A sound can be described as a function of time, $x(t)$. The value of the function can be the *amplitude* (in units of pressure) of the sound at a given point in space. Another useful value for the function is the *intensity* of the sound reaching a small area in space. A sound's intensity is the power transmitted through the area.

The human ear can hear an amazing range of sound intensities (or amplitudes). The intensity of an orchestra is about twenty millions times the intensity of a solo violin, but people hear both sounds easily because they perceive the intensity on a logarithmic scale. Thus, the intensity (or amplitude) of a sound is measured in *decibels* (dB), for it is a logarithmic unit. The Standard Pressure Level (SPL) decibel scale is defined as

$$\text{Sound level, dB SPL} = 10\log_{10}(I/I_0), \tag{1.1}$$

where I is the intensity level of the sound. I_0 is the reference level for the SPL scale: 2^{-16} W/cm^2. An equivalent form for an amplitude description is

$$\text{Sound level, dB SPL} = 20\log_{10}(A/A_0), \tag{1.2}$$

where A_0 is 20 μPa. The unit of measure for amplitude is *Pascals* and has units of Force/Area = Newton/meter^2 = N/m^2.

Another method of describing a sound that will be used often in this thesis is a *discrete-time function.* If a continuous-time sound is described by $x(t)$, then a discrete-time version would be defined as

$$\hat{x}[n] = x(t)\Big|_{t=\frac{n}{f_s}} \qquad n = 0, 1, 2, 3, \ldots , \tag{1.3}$$

where f_s is the *sampling frequency.* Thus a discrete-time function has the same value as the continuous function at specific points and is undefined in between these points. The value of f_s is a constant; therefore, the defined points in the discrete-time representation are evenly spaced in time. The discrete-time function can be thought of as a sequence of values.

A *digital* signal is defined as a subclass of discrete-time signals. If a discrete-time signal has integer values, then it is called a digital signal. The range of values that a digital signal can take is limited. The range is described in units of *bits.* An n-bit digital signal's range is

$$-2^{n-1} \leq x[n] \leq 2^{n-1} - 1, \tag{1.4}$$

where $x[n]$ is a digital signal. These signals are called "digital" signals because they are commonly used in digital electronics.

The process of mapping a continuous-time signal into a digital signal is called *sampling.* It is assumed that a continuous-time signal has continuous values; therefore, the integer-valued digital signal will have an error, $e[n]$, associated with every value. Mathematically,

$$x[n] + e[n] = x(\frac{n}{f_s}). \tag{1.5}$$

The value of $e[n]$ is

$$e[n] = x(\frac{n}{f_s}) - \text{int}\left(x(\frac{n}{f_s})\right), \tag{1.6}$$

where int() is a function that maps a continuous value into an integer value. If the digital signal's range is large enough (enough bits), then this error will be very small.

The reader may be confused as to how a digital function can represent a continuous time function. Nyquist proved (a good description of the proof can be found in Oppenheim and Schafer's book [OS89]) that a continuous-time function, $x(t)$, can be described by a digital function, $x[n]$, if $x(t)$ is *band limited.* The signal $x(t)$ is band limited if it has no frequency components at a higher frequency than one-half the sampling rate. This is called Nyquist's sampling theorem. In effect, Nyquist is saying that if the sound does not change significantly between samples (which means there are no high frequencies in the signal) it can be sampled. Now the frequency components of a signal can be discussed.

A frequency component of a sound is a sinusoid of a given frequency. Mathematically

$$A\cos(wt + \theta), \tag{1.7}$$

where w is the frequency of the component in radians, t stands for time, A is the magnitude, and θ is the phase. The frequency of a signal is more commonly measured in Hertz (Hz). The conversion from radians to Hertz is

$$2\pi f = w, \tag{1.8}$$

where f is the frequency in Hertz. Another method of specifying the frequency is to use the *period* or *wavelength* of the component. The period is defined as the inverse of the frequency and is measured in seconds.

Signals can be thought to be described by a summation of sinusoidal components. Such a description gives insight into the properties of a sound. For example, the pitch of a musical note is defined by its frequency components.

A means of determining the frequency components of a sound is required. If the signal $x(t)$ describes a sound, then the frequency components of the sound can be

determined using the Fourier Transform:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-jwt}dt, \tag{1.9}$$

where $e = 2.7182818284+$, the base of the natural logarithm, and $j = \sqrt{-1}$. $X(f)$ is a complex valued function that can be written as

$$X(f) = |X(f)|\, e^{j\angle X(f)}, \tag{1.10}$$

where $|X(f)|$ is called the *magnitude* of $X(f)$ and $\angle X(f)$ is called the *phase* of $X(f)$. The frequency components of a sound, called its *spectrum,* are plotted on two graphs, one for magnitude and the other for phase. A knowledgeable observer can better understand a sound by examining these plots.

The spectrum of equation 1.9 describes the average components of the sound over an infinite time. If the components in the signal are the same over the entire time of the signal, the signal is said to be in a steady state. Then the averaging effect of the Fourier Transform is unimportant. For this reason, a frequency spectrum derived from the Fourier Transform is called a *steady state frequency spectrum.*

Now that frequency components have been described, Nyquist's sampling theorem can be more rigorously defined. Mathematically speaking, a continuous-time signal can be represented by a digital (or discrete-time) signal if and only if the continuous signal is band limited. A signal $x(t)$ is band limited if and only if

$$X(f) = 0 \qquad \text{for } f \geq f_s/2, \tag{1.11}$$

where $X(f)$ is the Fourier Transform of $x(t)$. Thus, if a sound contains no frequency components above half the sampling frequency, it can be completely represented by a sampled version of itself.

The frequency components of a digital (or discrete-time) signal can be determined using the Discrete Fourier Transform (DFT). The steady state frequency spectrum

of a sampled signal, $y[n]$, is defined as

$$Y(e^{j\Omega T}) = \sum_{n=-\infty}^{\infty} y[n]e^{-j\Omega Tn}. \tag{1.12}$$

As the function name $Y(e^{j\Omega T})$ implies, the function is periodic. This is the effect on the signals spectrum caused by sampling. The variable ΩT is related to the frequency of the signal by the equation

$$\Omega T = \frac{2\pi f}{f_s}. \tag{1.13}$$

A great deal of information can be determined from the spectrum of a sound. Figure 1.1 shows the magnitude of the frequency components for several sounds plotted on a logarithmic frequency scale. Each spectrum can be described in general terms. These terms are usually used to describe musical signals, but herein they will be used to describe non-musical signals as well. The sound spectrum in part **a** has predominantly high frequency components. Thus it is said to sound *trebly*, or have a lot of *treble*. The sound spectrum in part **b** has strong low frequency components, thus it has a deep *bass* tone. The sound spectrum in part **c** has strong mid-frequency components. This means it has a tinny or neutral tone. This is generally undesired because the treble and bass in a sound give it its flavor. If the peak is very sharp and narrow, the sound is said to be *narrow-band*. This means only a narrow set (or band) of frequency components are present in the sound. The spectrum in part **d** is called *white noise*. The term "white" means that the sound contains all frequency components, just like white light contains all wavelengths of light. *Coloured* noise contains only certain bands of frequencies, similar to coloured light.

In summary, a definition of sound has been established. How sound is described as a digital signal, how the spectrum of a sound is determined and the properties of a sound's spectrum help to define a common vocabulary for sounds. Systems are discussed next.

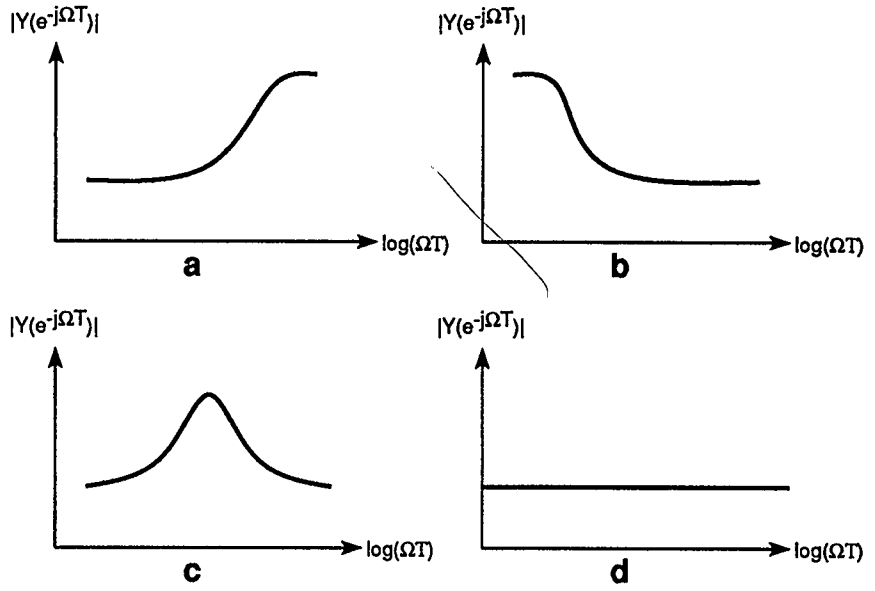

Figure 1.1: Magnitude spectrums of some sampled sounds.

1.2 Systems, Digital Systems, and Frequency Response

What is a system? A system is a black box with inputs and outputs. The output of a system Q when the input is $x(t)$ is written as $Q\{x(t)\}$.

A digital system is a system whose inputs and outputs are digital signals. In this thesis only linear, time-invariant, causal systems are considered.

Given two input functions, $x_1[n]$ and $x_2[n]$—and corresponding outputs from the system, $y_1[n]$ and $y_2[n]$—then a system is linear if and only if

$$Q\{ax_1[n] + bx_2[n]\} = aQ\{x_1[n]\} + bQ\{x_2[n]\} = ay_1[n] + by_2[n], \tag{1.14}$$

where a and b are constants (Note that if the digital signal's range is insufficient to represent the multiplication values, a non-linearity will occur. This topic is discussed in Oppenheim and Schafer's [OS89] book). If a system is causal, then there is no output before an input. In other words, an output is *caused* by an input. Mathematically, if $x_1(t) = 0$ when $t < t_0$, then $y_1(t) = 0$ for $t < t_0$. A system is time-invariant

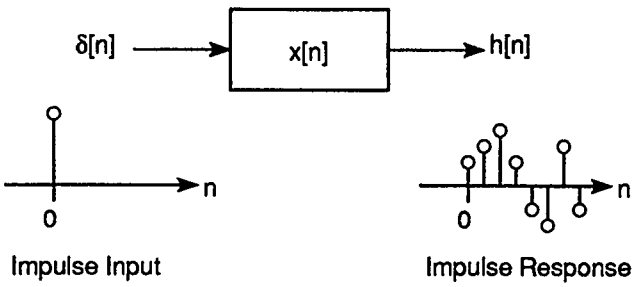

Figure 1.2: The impulse response of a system.

if the same input at a different time provokes the same output. More precisely, a system with input $x_1[n]$ and output $y_1[n]$ is time invariant if and only if

$$Q\{x_1[n+n_0]\} = y_1[n+n_0] \qquad \text{for all } n_0. \tag{1.15}$$

A system that is linear, time-invariant, and causal is called a Linear Time-Invariant (LTI) system.

An LTI system can be completely described by its *impulse response.* If the input to a system is a unit sample, defined as

$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise,} \end{cases} \tag{1.16}$$

the output of the system, $h[n]$, is the impulse response of $Q\{\}$. This is depicted in Figure 1.2. If $h[n]$ is the impulse response of Q, $x[n]$ is the input, and $y[n]$ is the output,

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] \tag{1.17}$$

because the system is linear, causal, and time-invariant. The previous equation is called the *convolution summation.* A digital LTI system is completely defined by its impulse response because its output can be determined from the impulse response and the input function.

The digital systems used in this thesis are *filters* with one-input and one-output. Digital filters are defined as any system that can be described by *difference equations* of the form

$$y[n] = \sum_{k=0}^{M} b_m x[n-m] - \sum_{k=1}^{N} a_k y[n-k], \tag{1.18}$$

where a_k and b_m are constants.

It is important to determine how the output of a filter is related to the input. The impulse response is one input/output relation, but it is better to describe the relation in regards to the frequency domain. The frequency domain relation is called the *frequency response* of the system. The frequency response can be determined from the impulse response using the Discrete Fourier transform;

$$H(e^{j\Omega}) = \sum_{k=-\infty}^{\infty} h[k]e^{-j\Omega k}, \tag{1.19}$$

where $H(e^{j\Omega})$ is the steady state frequency response of the system. It can be proven that convolution of the input with the impulse response is equivalent to multiplication of the Fourier transform of the input by the frequency response. Thus, if $\mathcal{F}\{\}$ is the Fourier transform,

$$H(e^{j\Omega})\mathcal{F}\{x[n]\} = H(e^{j\Omega})X(e^{j\Omega}) = Y(e^{j\Omega}) = \mathcal{F}\{y[n]\}. \tag{1.20}$$

The output frequency spectrum, $Y(e^{j\Omega})$, is the product of the input spectrum, $X(e^{j\Omega})$, and the frequency response of the system, $H(e^{j\Omega})$. When $H(e^{j\Omega})$ is used in this manner, it is called the *transfer function* of the system because it "transforms" the input into the output.

An LTI system's frequency response describes how a sound will be modified by the system. Each graph in Figure 1.1 is the frequency response of a system, and the systems can be classified according to how they alter a sound. The response in part **a** is that of a *hi-pass* system because it lets high frequencies pass through (or

boosts them) while attenuating low frequencies. It would give a sound more treble. The sound in part **b** is called a *low-pass* system and would give a sound more bass. The response in part **c** is called a *band-pass* response. It is useful in boosting or isolating a specific band of frequencies in a sound. Finally, the response in part **d** is called an *all-pass* response because it passes all frequencies equally. This may seem useless, but only the magnitude response has been specified. The phase response of the system may be altered.

The important concepts of this section are digital systems, their frequency responses and their impulse responses. These concepts not only contribute to a common vocabulary, but are referred to throughout this thesis. Another way of looking at digital systems which enhances this section is discussed next.

1.2.1 The z-transform

The z-transform, $\mathcal{Z}\{\}$, is a convenient way of representing a digital system. Using the nomenclature and definitions from Oppenheim and Schafer's book [OS89], the z-transform is defined as

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}, \tag{1.21}$$

where z is a complex variable and $X(z)$ is the transform of $x[n]$. The summation does not have to converge for all values of z. For this reason, a region of convergence for the function should be stated as

$$r_- < |z| < r_+, \tag{1.22}$$

where r_- and r_+ are constants that define the region of values that z can take.

Multiplication in the z-domain is equivalent to convolution in the digital-domain. If $X(z) = \mathcal{Z}\{x[n]\}$, $Y(z) = \mathcal{Z}\{y[n]\}$, and $H(z) = \mathcal{Z}\{h[n]\}$, then

$$Y(z) = H(z)X(z) = \mathcal{Z}\left\{ \sum_{k=-\infty}^{\infty} x[k]h[n-k], \right\}. \tag{1.23}$$

The z-transform of a digital system's impulse response, $H(z)$, is the z-domain transfer function. The transfer function of a system can also be determined in the z-domain. If $Y(z)$ is the output of a system and $X(z)$ is the input, then

$$H(z) = \frac{Y(z)}{X(z)}. \tag{1.24}$$

It can be shown that the steady-state, sinusoidal, frequency response of a system can be determined from its z-transform by substituting $z \rightarrow e^{j2\pi f/f_s}$.

Other useful properties of the z-transform are the *unit delay* and *linearity* properties. The unit delay property is described as

$$\mathcal{Z}\{x[n-k]\} = z^{-k}X(z). \tag{1.25}$$

Thus a delay of k samples is the same as a multiplication in the z-domain by z^{-k}. z^{-1} is called the unit delay operator. Linearity states that if

$$w[n] = ax[n] + by[n],$$

then

$$W(z) = aX(z) + bY(z). \tag{1.26}$$

These two properties make it easy to determine the z-domain transfer function of arbitrary networks consisting of unit delays, additions, and fixed-coefficient multiplications. Delay, addition, and multiplication are the basic building blocks of digital signal processing.

For completeness, and to show an example of a digital system, a graphical notation for describing systems consisting of these three building blocks is introduced. The graphical notation consists of nodes and branches. Branches are represented by arrows and nodes are where branches meet. Every node in a graph has a numeric value. The branches describe how the node's values are related. The value of a branch is equal to the value of the node at the tail end multiplied by the value

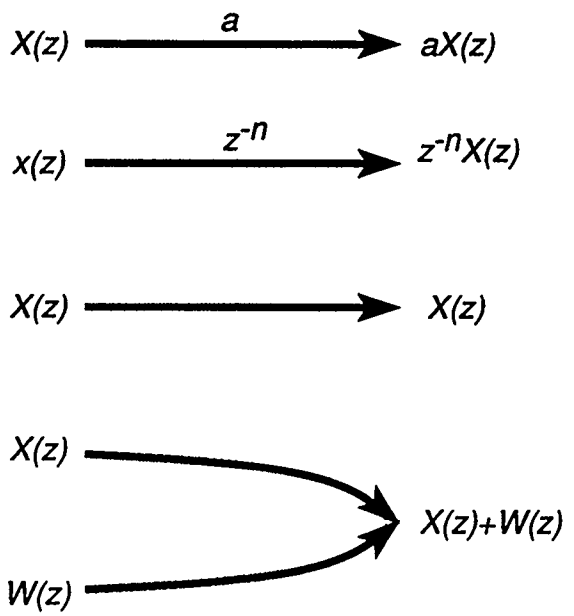

Figure 1.3: Discrete network elements.

written next to the branch. If no multiplier is specified, it is assumed to be unity. The value of a node at the head of a branch is equal to summation of the values of all the branches that point to it. Figure 1.3 gives a graphical description of these specifications and Figure 1.4 specifies a network using the notation. The z-transform of the system in Figure 1.4 can be determined by solving the equations presented in the graph:

$$\begin{aligned} Y(z) &= X(z) + aY(z)z^{-1} \\ Y(z)\left(1 - az^{-1}\right) &= X(z) \\ \frac{Y(z)}{X(z)} &= H(z) = \frac{1}{1 - az^{-1}}. \end{aligned} \tag{1.27}$$

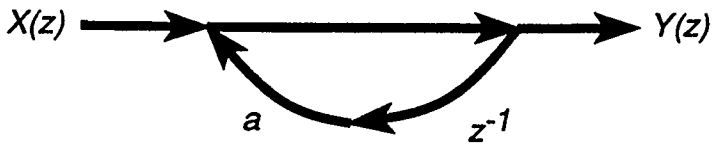

Figure 1.4: An example network.

1.3 Hearing: A Sensory Perception

People live in an *environment.* An environment is the objects and circumstances surrounding a person. This includes everything real to the physical world, such as air, rocks, electro-magnetic radiation, velocity, etc. People interact with their environment by picking up things, moving things, eating things, inhaling things, etc. It is necessary for people to interact with their environment in order to survive.

To productively interact with the environment, people must perceive their environment. For example, when people are hungry, they have to be able to find food. People perceive their surroundings using five *senses*: touch, taste, smell, vision and hearing.

What does it mean to perceive something? Perception requires two elements. Blauert [Bla83] states there is no perception without a *subject* and an *object.* The object is perceived, the subject perceives it. Herein the subject will always be a person. An object is less easy to define. It is any physical thing (a rock), or a feature of any physical thing (the colour of a rock). Thus, perceiving is the act of a subject becoming aware of an object.

How does the subject perceive the object? The object alters its environment. For example, a drum agitates the surrounding air when struck. The alteration of the environment is transmitted to the subject by a *medium.* The subject has a sensory organ that reacts to changes in the medium. The sensory organ sends information

about the change to the brain via a series of nerve impulses. The brain interprets these nerve impulses to have some meaning and the process is complete. Note that perception is defined to be not only the process of receiving information, but also the interpretation of it.

Hearing is the sense investigated in this thesis. A subject *hears* when they notice a sound and gain information about an object from it. This is a definition of hearing.

The intricacies of human hearing must be discussed before proceeding. Most people understand the intricacies of sight in relation to depth perception, focusing, edge enhancement, and other phenomenon. However, most people know relatively little about similar phenomenon in hearing. The next few sections familiarize the reader with hearing perception.

1.4 Loudness

Loudness is the perceived strength of a sound signal. One would think that the more sound intensity, the louder the sound. But this is only a general trend and many exceptions exist to this rule. For example, the ear will adapt to loud sounds by lowering its sensitivity; thus, the perceived loudness can not be predicted unless the subject's immediate listening history is known.

Another complication is that the ear is not equally sensitive to all frequencies. The ear has a specific frequency response. Above 20kHz and below 20Hz the ear is not sensitive at all. In between these two limits, the ears sensitivity fluctuates with the general peak of sensitivity at 1kHz, and the sensitivity droping off in both directions. The ear's sensitivity to various frequencies is dependent on the sound intensity, and the sensitivity is different for every individual. In general, the more intense the sound, the better the bass and treble response of the ear.

Since the ear is not sensitive to all frequencies equally, the frequency spectrum of

a sound should be interpreted with caution. For example, if the higher frequencies of a sound are boosted, the effect will be less noticeable than a boost to frequencies around 1kHz. As a result, the perceived loudness of a sound is not easily determined. It is a complex quantity that depends on the frequency spectrum and intensity of the sound. Physical measurements of signal intensity are only guidelines to the perceived loudness of a signal.

1.5 Pitch

Musical pitch is briefly discussed because it shows the frequency perception of the human ear is logarithmic.

The pitch of a sound is defined by its frequency components. Sinusoids are a pure pitch. They have musical meaning. For example, a 440Hz sinusoid is a concert tuning A-natural note. An *octave* is a common interval for musical notes. For example the note one octave above concert A-natural is a 880Hz A-natural note. One octave above that is a 1760Hz A-natural note. Going up one octave doubles the frequency of the sinusoid. To the human ear the interval sounds equal. The 440Hz jump in the first example and the 880Hz jump in the second example sound the same. Thus the human ear perceives pitch in a logarithmic manner (octaves). For this reason, the spectrum of sounds and the frequency response of systems will be plotted on a logarithmic frequency axis.

1.6 Anatomy of the Human Ear

Another important aspect of hearing is the actual shape of the human ear. The fleshy external part of the ear is shown in Figure 1.5. This part of the ear, as a whole, is called the *pinna*. All the features shown in the diagram are present in any human pinna, but the proportions of the elements vary greatly. A pinna is as unique

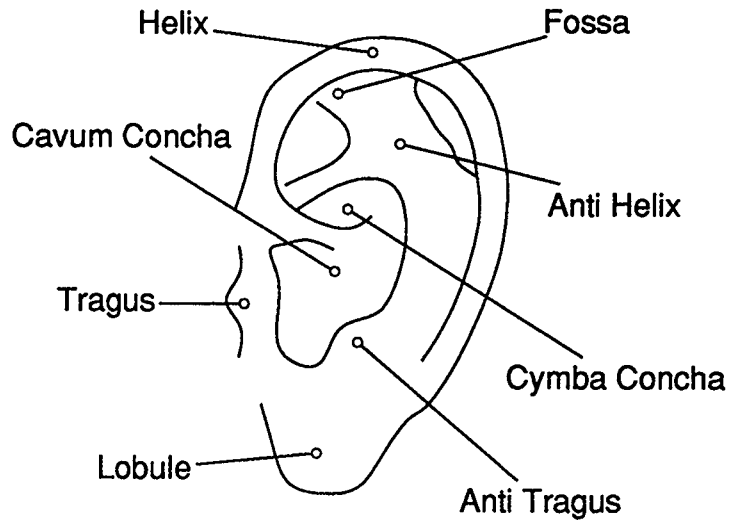

Figure 1.5: The pinna.

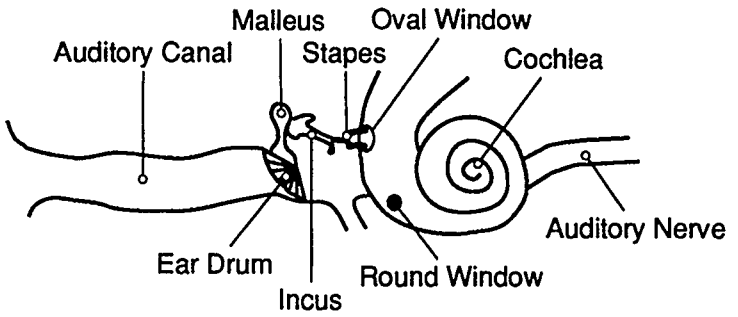

Figure 1.6: The ear canal, middle ear and cochlea.

to an individual as their fingerprints.

The pinna will alter the sound reaching the ear by reflection and refraction. These alterations are dependent on the direction of the sound and are useful in the localization of a sound. This is discussed in greater detail in section 1.8.

The remaining parts of the ear are shown in Figure 1.6. The pinna, ear canal and ear drum are referred to collectively as the external ear. The pinna acts together with the ear canal to focus and shape the sound reaching the ear drum.

The middle ear transfers sound vibrations from the ear drum to the cochlea. It

consists mainly of three bones: the malleus, incus and stapes. The purpose of the middle ear is to act as an impedance matching device between air and the fluids of the cochlea. The ear fluids are more resistant to sound vibrations than air is. As a result, sound waves would mostly reflect off the resistant ear fluids, but the middle ear is roughly the same resistance as air on one side and the same resistance as the ear fluid on the other. By matching the impedance, sound is transferred efficiently to the cochlea.

The cochlea is the most important part of the ear. It translates the vibrations of sound into nerve impulses that are transmitted to the brain. Lyon and Mead [LM88] describe the cochlea as a liquid filled tube. The tube is lined with hairs which are divided into two groups: inner hairs and outer hairs. The inner hairs are connected to neurons that signal to the brain that sound is present. A signal is sent when movement in the cochlea fluid causes an inner hair to be bent.

The middle ear pushes the liquid in the cochlea by exerting pressure on the oval window (see Figure 1.5). The round window allows the fluid somewhere to flow. The fluid dynamics of the cochlea are such that a frequency selection takes place. The fluid pressure against an inner hair is a bandpass function. Thus different hairs respond to different frequency bands. This means that the cochlea performs a crude frequency analysis before passing information to the brain. The cochlea does not perform a Fourier transform, but information about the magnitude spectrum of a sound is present. Lyon and Mead [LM88] state that the cochlea only provides information about frequency bands, not a rigorous spectrum analysis of a sound.

Moore [Moo82] reports a strange phenomenon that shows the brain receives information about the phase of a sound's spectrum also. If a pure sinusoidal tone is played into an ear, the firing of the neurons connected to the inner hairs will be phase-locked with the sinusoid. In other words, the neuron will fire at a specific

point on the sinusoid. The neuron firings also lock onto the period of sounds more complex than sinusoids.

If the outer hairs of the cochlea are not connected to neurons, what is their purpose? The outer hairs provide positive (or negative) feedback to the sound input to the ear. The brain controls the behavior of the hairs so as to provide attenuation and gain to signals. Thus the brain has some control over what reaches the ears neurons. The outer hairs reinforce quiet sounds, making them more audible. This feedback mechanism can even make the ear ring as the system becomes unstable. Zurek and Clark [ZC81] observed that chinchilla ears could resonate and produce a squeal audible up to three meters away. Thus the ear has a means of automatic gain control that it uses to adjust the volume of the incoming sounds. It is similar to the dilation of the eye to adjust the level of light entering the eye. It is suggested that this also allows the listener to "focus" on certain parts of a sound. This suggests that any measure of loudness is relative to the person who is hearing it and how much they are attenuating/boosting the incoming signals. Thus, when setting up an auditory experiment the concentration level of the listener is a factor.

It should be noted that the neuron signals reaching the brain are probably processed by a dedicated part of the brain in some predesigned manner. This would be similar to the way the eye signals are processed to enhance edges and complete lines. The brain probably encodes the neuron firings so as to highlight certain traits of the sound. What traits are enhanced is largely unknown, but one example is the precedence effect.

Zurek [Zur87] describes this phenomenon in the following manner. If two identical sounds are heard, one slightly before the other, the first incident sound will be reinforced. The auditory system automatically tries to cancel out echos to make a sound clearer. If not for this capability, the echos in an environment would make it

impossible to carry on a conversation.

The human ear has been shown to be a complex system. Some insight into the hearing process has been gained by examining the mechanical operations of the ear. For example, the brain is given information about the frequency components of a sound, both magnitude and phase. Also the ear has a gain control system based on feedback that allows it to control its input (from the neurons). It is these intricacies of the ear that are vital to many areas discussed throughout this thesis.

1.7 Interpretation of a Sound

The way that the brain interprets sounds is a psychoacoustic topic. It involves psychological elements as well as acoustic. When a subject hears a sound, they assign attributes to it and use it to describe an object. How does the brain do this?

The brain has a stored set of previously heard sounds. When it hears these sounds again, it will assume that the same object is making the sound. The information about the object that is stored in the subject's memory can be recalled and used. If the sound is not in the stored memory, it will be learned and added.

A sound is altered by the environment that it is heard in which can adversely effect pattern recognition. For example, walls reflect sound. Fortunately the human auditory system will adjust to the environment it is in. It can learn and memorize listening environments and then "un-distort" a sound heard. For example, if the sound is being reflected, the first arriving incident of the sound will be enhanced. This is the *precedence effect* as described by Zurek [Zur87].

When a sound is heard the human auditory system attempts to give it a location. In some people, sound localization is very developed. When they hear a sound, they instantly know where it is located. Sound localization is beneficial because people can not always visually locate objects in their environment. Their ears become the

prime information source for objects not in their visual path. This is apparent for blind people who have to develop their hearing to a great degree, but some people never develop their hearing and rely mostly on sight for localization. Those who cannot localize sound well are called *poor localizers.*

It should be emphasized that hearing does not work alone. People use it in conjunction with the other senses and higher brain functions. All sensory data available is used to make a complete picture of the environment. For example, if a person hears a feline growl in Africa, they know it isn't a tiger. Tigers don't live in Africa. A higher brain function has effected the pattern recognition of an object. When people see a ventriloquist with a puppet on their knee, their localization is fooled. When the puppet's mouth moves and the ventriloquist's doesn't, they assume the puppet is talking and not the ventriloquist. Visual data is overriding the hearing localization.

In summary, interpretation of a sound is a complex and difficult to predict function of the brain. It depends on the stored experience of the listener, the current environment, the object location, and the other senses.

1.8 Localization

When a subject hears a sound, they assign several properties to it. They will try to identify what object made the sound; is it a lion's roar or a bubbling brook. At the same time the brain will try to determine where the sound is coming from. This is called sound localization.

A careful distinction between the physical and perceived locations of a sound must be made. The location of an object (the source of the sound) in an environment is a physical quantity. There is also the location of the object in the subject's auditory environment or where the subject thinks the sound is. The two locations do not have

to correspond; thus, a distinction between the two can be made. The actual location will be referred to simply as the *event's* location. If it is necessary to distinguish that the location is the perceived location, it will be referred to as the *auditory* location. Whenever the adjective "auditory" is used, the perceived event is being referred to.

An example will help clarify the difference between an actual and auditory event. Assume there are two speakers placed a few meters apart from each other and an equal distance from the listener. Let the two speakers project a sound of equal intensity simultaneously. There are two events: one at the left speaker and one at the right speaker. These are the actual physical events. In such a situation there will be only one auditory event. Its location will be directly between the two speakers. The auditory location would have the sound emanating from a phantom sound source. This phenomenon is called *summing localization.*

1.8.1 The Auditory Environment

Summing localization is one of the ways in which the real environment and the auditory environment can differ. Summing localization occurs when two similar sound events are close together. If the two events are of equal loudness, then the auditory event will be located between the two sources. If one event is louder than the other, the auditory event will shift towards the louder signal. Thus, it is possible to use two sound sources in the real environment to create an event anywhere between the two sources in the auditory environment. This is the basis of stereo sound technology. Further information about summing localization can be obtained from Blauert's book [Bla83].

Sounds in the real environment are located more precisely than in the auditory environment. Auditory objects are located with reference to the listener, so the polar coordinate system shown in Figure 1.7 is used (zero degrees is straight ahead of

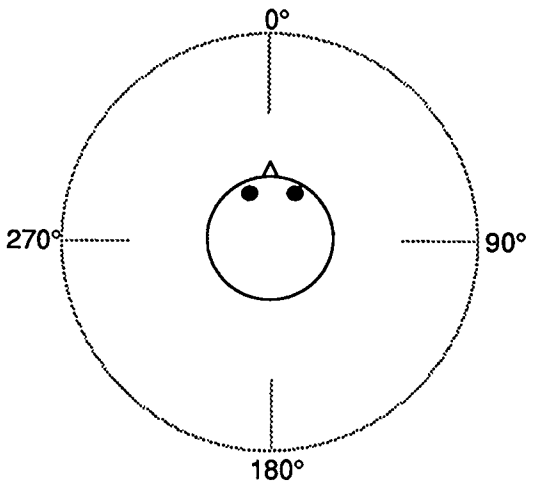

Figure 1.7: Coordinate system for auditory space.

the listener and the angle is measured clockwise from zero). The auditory objects location will be *blurred* with respect to the actual location. There are two types of auditory blur: distance and angle blur. Blauert [Bla83] notes that distance localization is largely based on the relative loudness of the sound. The subject must be familiar with the sound before its loudness can be used to determine its distance. Thus, the distance blur is dependent on learning. Haustein [Hau69] showed that for a familiar sound, the localization blur is dependent on the distance from the subject. Distance blur for a single point source can be defined as $\Delta r(r, l)$, where r is the distance of the object from the subject and l is a learning parameter. In general, $\Delta r(r, l) \approx 0.5$m when the listener is in a normal room, but this approximation will change with the distance and familiarity of the sound source.

The angular blur for a single point source is represented by the function $\Delta\theta(\theta, l)$. It is dependent on the direction of the incident sound and previous knowledge about the sound. Blauert [Bla83] surveys many papers that investigate angular blur. He shows that $\Delta\theta(\theta, l) \approx 2°$ when the listener is in a normal room, but this blur will change with the angle of incidence and familiarity of the sound source. If distance

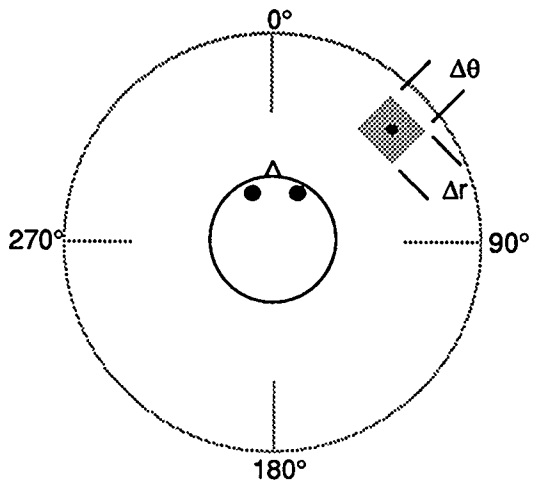

Figure 1.8: A blurred point in auditory space.

and angle blur are both taken into account, a point in real space can be considered an area in auditory space as shown in Figure 1.8.

One final anomaly of the auditory space is discussed. Sometimes the perceived event is located at the mirror image position (reflected about a line through both ears) of the actual event. For example, a sound at 30° is localized at 150°. This is called *front-back reversal.* Blauert [Bla83] makes two observations about this phenomenon. First, front-back reversal can be eliminated by head movement because moving the head gives much more information about a sound. Head movement is useful in many situations where localization is difficult or erroneous. Second, narrow-band signals or "unnaturally" altered signals are often reversed. This suggests that the frequency components of a signal effect localization.

In summary, there are two different sound environments: the real and auditory environments. The real environment is the one people live in. The auditory environment is the one people perceive through hearing. The auditory environment is blurred and inexact. Phantom sources caused by summing localization and front-back reversals are common.

1.8.2 How Sound Is Localized

The *Duplex Theory* is the original theory of localization developed by Lord Rayleigh [Ray07]. It states that there are two cues for localization. The first cue is the time difference between a sound reaching each ear. If a sound source is closer to the right ear than the left ear, the sound must travel further to get to the left ear. The time it takes the sound to travel the extra distance is a cue to the location of the sound. This cue is called Interaural Time Difference (ITD). The further a sound must travel, the lower its signal power will be. If it must travel around an object, such as the human head, its signal will be further attenuated. Thus, the volume of the sound will be louder at one ear. This localization cue is called Interaural Intensity Difference (IID).

Lord Rayleigh hypothesized that low frequency signals, with their large wavelengths, do not get attenuated by going around the head because of their long wavelength. He concluded that the IID must be a poor cue for low frequency signals and used mainly for high frequency localization. In addition, it can be seen that if the period of a signal is shorter than its interaural time delay, then the cue can't be detected. The subject can't tell how many periods have elapsed. Thus, the ITD cues become meaningless at high frequencies. Rayleigh concluded that the ITD cues must be used primarily to locate low frequency signals.

Summing localization and other auditory phenomenon are explained by the duplex theory. But the duplex theory can not account for localization along the *cone of confusion* shown in Figure 1.9. If the head is assumed to be a perfect sphere and the ears (represented by two points connected by a line in the figure) are placed on an axis through the center of the sphere, there are several object locations with the same interaural cues. These locations define the cone of confusion. Note that the mirror image of a sound in front of a listener is behind the listener on the cone of

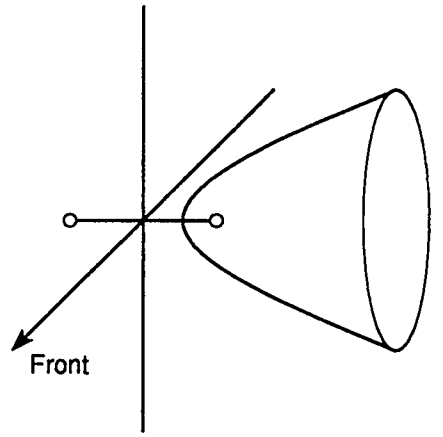

Figure 1.9: Cone of confusion for interaural localization.

confusion. This explains front-back reversals. Batteau [Bat67] states localization is still possible when the listener is completely deaf in one ear. The Duplex theory can not explain localization with one ear or along the cone of confusion, so it is not a complete explanation of localization.

Originally it was believed that sounds along the cone of confusion were localized using head movement. But Fisher and Freedman [FF68] showed that people can localize sounds when their heads are completely restrained. Thus some other cue must be at work.

Blauert's paper on sound localization [Bla70] shows that the pinna provides localization cues. He experimented with sounds in the median plane. The median plane, a degenerative form of the cone of confusion, is composed of all points an equal distance from each ear. Thus, all objects in the median plane have the same interaural characteristics. Blauert discovered that the location in the median plane (either forward, backward, or above) is not determined by the direction of incidence of the sound, but by the frequency components of the sound. He showed that the pinna alter the frequency spectrum of a sound depending on the direction of incidence, and this is confirmed by Shaw and Teranishi [ST68].

One of the main arguments against localization by spectral modification is that it would require previous knowledge about the spectrum of the sound. People can locate sounds they have never heard before, but if a subject does not know the original sound spectrum, how can they guess what has been altered? Blauert suggests that people localize by perceiving which frequency bands in a sound are most prominent, not which bands have been altered. Certain frequency bands, called *localization bands,* are associated with certain locations. If the most prominent frequencies in a sound are within a localization band, then the sound will be localized at that band's location. If this theory is correct, it should be possible to change the location of an auditory object by artificially boosting some of its frequencies. Blauert proves this in an experiment by showing that band limited noise has a fixed auditory location even if the physical location is changed. Thus, the frequency spectrum alterations caused by the pinna are used to localize objects on the cone of confusion. Fisher and Freedman [FF67] showed that single-ear localization using pinna cues can be as accurate as binaural localization, and front-back reversals are much less common when only pinna cues are used.

Thus, when the head is not moving, there are two complementary localization systems working together. The first uses binaural cues such as IID and ITD, but front-back reversals will occur because interaural cues can not resolve locations along the cone of confusion. The second system uses pinna cues and is less susceptible to front-back reversals. The sounds frequency spectrum is altered by the pinna depending on its angle of incidence, and the altered spectrum provides cues for localization.

Fisher and Freedman [FF68] showed experimentally that pinna cues are required in hearing. They performed tests with head movement allowed and head movement restricted. In the first test, the subjects listened normally. Then they placed ten

centimeter long tubes in the subject's ears to nullify the effect of the pinna. Finally, they placed artificial pinna on the end of the tubes. When head movement was allowed, all three cases performed similarly because head movement is a good source of localization cues. With the head restrained, the accuracy of listening with pinna (their own or the artificial pinna) was much better than listening without pinna. Thus, it can be concluded that pinna cues are very important to still-head localization. This experiment also proved that artificial pinna can give adequate cues for localization. In other words, people can hear through pinna other than their own.

1.9 Headphones and Soundstages

Any listening environment that uses two speakers very close to the ears is referred to as headphone listening. Headphones have a peculiar *sound stage.* A sound stage is defined as the total area from which sound can originate in a listening environment. In a typical stereo environment (stereo being two channel recordings; left and right) speakers are used. The speakers are placed 30° to the left and right of the listener as shown in Figure 1.10. The sound stage in such an environment, known as a *free-field environment*, is a straight line between the two speakers. This implies that in a free-field listening environment, an auditory event can be located at any position between the two speakers. This is the environment in which most recordings are meant to be listened to. Sound systems in movie theaters have six speakers, as shown in Figure 1.11. The sound stage is anywhere within the rectangle enclosed by the speakers.

The sound stage for headphones is a straight line between the two earphones (see Figure 1.12). Thus the sound stage passes through the head. This is a peculiar and unnatural effect. The only sound people hear inside their heads is the grinding of their teeth. Bauer [Bau61] gives an explanation of why headphones give a poor

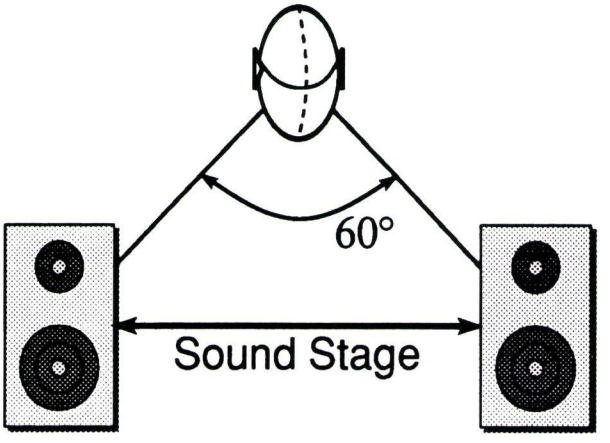

Figure 1.10: Sound stage for free field listening.

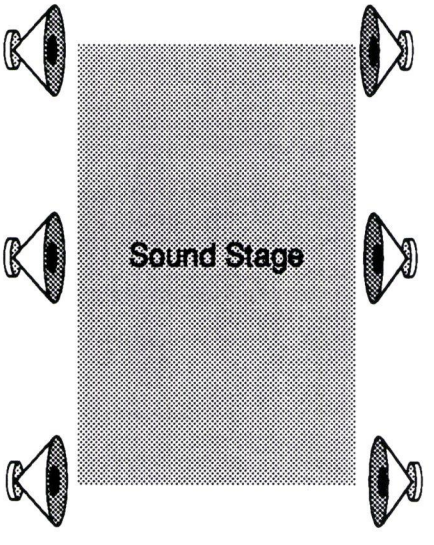

Figure 1.11: The two-dimensional sound stage in movie theaters.

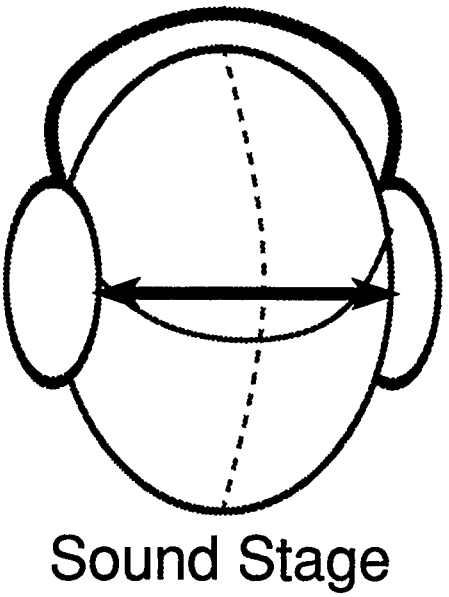

Figure 1.12: Sound stage for headphones.

stereo image. If a recording is made for listening over speakers, a sound may be recorded on just the left side of the recording. If this is listened to on speakers the sound will reach both ears. This is natural. When the same recording is listened to on headphones, the sound is only heard in the left ear. In a real environment this would only occur if the sound source was very close to that ear and very quiet (so as not to be overheard by the other ear). If the program is listened to at a loud volume, the sensation has no analogy in nature and so causes confusion. Bauer suggests that when the auditory system is confused, it localizes sound inside the head.

Headphones promote *lateralization* instead of localization. Plenge [Ple74] defines lateralization as the localization of a sound inside the head. It should be noted that it is possible to enhance the sound stage in headphone listening so that we can localize rather than lateralize. That is the first objective of this thesis. Conversely, lateralization is also possible in free-field listening. In lateralization the subject perceives Inside-the-Head Localization (IHL) and in localization the subject perceives

Outside-the-Head Localization (OHL).

Headphones provide an imperfect listening environment, but they provide a means of individually controlling the input to each ear and they are not effected by the surrounding environment. Speakers do not control the input to each ear individually because there is cross-talk. Cross talk means that the sound from the right speaker will reach the right ear and the left ear[1]. In headphone listening, the right channel of a stereo recording is played directly into the right ear and does not effect the left ear. In speaker listening, reflections from objects in the environment and environmental noise will interfere with the sound heard. In headphone listening the environment is bypassed and these problems are eliminated.

1.10 Objective

This thesis will discuss a method of controlling the localization of auditory events using headphones. This method will be used to achieve two objectives. The first objective is to use headphones to simulate a free-field sound stage. The second objective is to simulate multi-channel environments using headphones.

Multi-channel recordings (when this thesis refers to multi-channel recording it will mean more than two channels) have become popular in motion pictures because they provide a two dimensional sound stage. A channel is a recorded source meant to be listened to on a separate speaker in a free-field environment. A multi-channel system places greater demands on the recording medium than a typical stereo recording. Humans only have two ears, and so they can only hear two signals. Thus, it should be possible to represent a multi-channel environment using only two signals presented over headphones. This is a form of data compression using the ability of the brain

[1]It is possible to control the input to each ear individually using speakers. If the sound from the near speaker cancels out the sound reaching the ear from the far speaker exactly then cross-talk is eliminated and the ears are isolated. This is a difficult procedure. The environment the speakers are in must be compensated for and the listener must remain motionless.

to interpret directions to get more information in a limited channel. The brain is a powerful receiver if a method of transmitting to it can be devised.

The first objective is a simplification of the second objective. A free-field listening environment is a multi-channel environment with only two channels. It is proposed that if a free-field system is simulated, the method of simulation can be extended to multi-channel environments.

What method should be used to localize the sounds?

Head movement cues are one method of stimulating localization in the listener. This method simulates the IIDs and ITDs for every orientation of the head. If the head moves through an arc, the IIDs and ITDs will change. Loomis, Hebert, and Cincinelli [LHC90] proved this information allows the listener to localize, but this method will not be used because it is inappropriate. People do not normally move there heads while watching movies or listening to recordings, and the head movement method can not simulate an environment if the listener's head is stationary.

If the listener's head is stationary, a combination of IID, ITD, and pinna cues should be used. The following chapters document various attempts at controlling the location of auditory objects using these cues.

Chapter 2

The Echo Model

The first attempt at simulating localization cues was based on a paper by Batteau [Bat67]. Using measurements on human ear replicas that were five times human size for convenience, he showed an echo is present in an ear signal, and the echo time is dependent on the angle of incidence in a linear fashion. Hence, the ear might determine localization by analyzing the echo time in a sound.

This may seem to contradict Blauert's theory of localization frequency-bands, but the two theories complement each other. If an echo is present in a signal it can be modeled by the digital comb filter in Figure 2.1. The filter's frequency response, shown in Figure 2.2 for $m = 4$ and $f_s = 44100$, has a number of boosted and attenuated frequency bands. Thus, an echo model can account for the boosted frequency bands in Blauert's theory.

2.1 Definition of an Echo Model

The echo model is based on the features of the pinna. Batteau indicates that the main features of the pinna are two bowl like surfaces. These are the cavum concha and the helix of the ear. Figure 2.3 shows the location of these two reflectors. In theory, these surfaces add cues to the sound in the form of an echo. Reflector two provides cues for vertical location of a sound and reflector one provides cues for

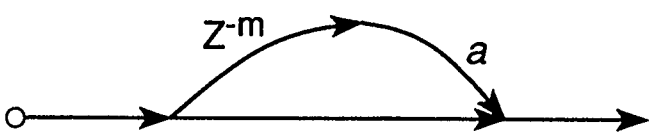

Figure 2.1: A comb filter.

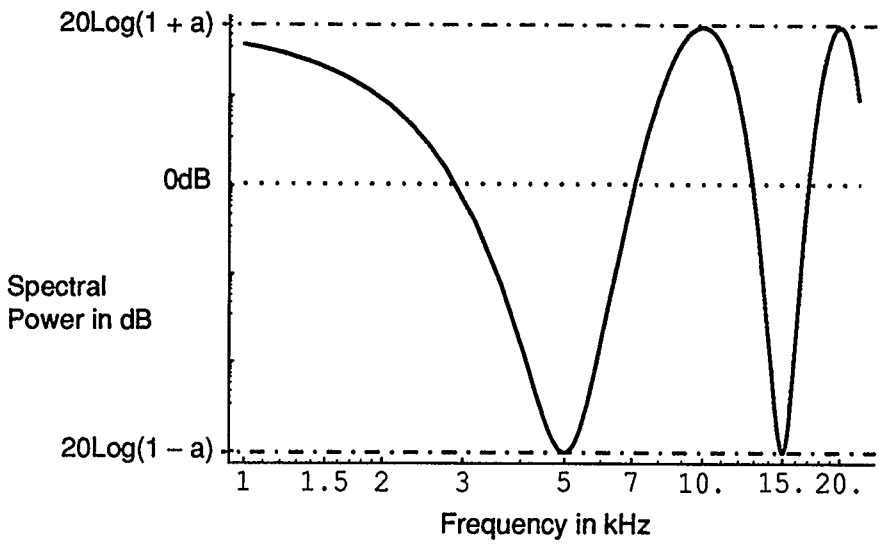

Figure 2.2: Magnitude, frequency response of a comb filter.

horizontal localization. The reflectors shown only account for sounds from the front and below the ear, but similar reflectors can be constructed for sound from the rear and above the ear.

A major criticism of this theory is that it requires the brain to detect extremely small echo times; ranging from approximately $10\mu sec$ to $300\mu sec$. Wright, Hebrank, and Wilson [WHW74] showed that the human ear is capable of resolving these very short echoes. As was shown, an echo will boost and attenuate frequency bands in the signal, and the brain might perceive the echo by perceiving these spectrum distortions.

2.2 Simulating the Echo Model

The echo model from Batteau's paper can be simulated using a digital computer. Digital audio systems in this thesis were simulated using a NeXTstation computer running the NeXTStep 3.0 operating system, and sound was sampled using a Sony

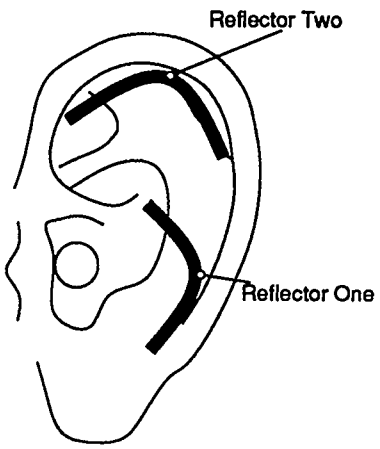

Figure 2.3: Reflection surfaces on pinna that lead to a echoed signal.

DTC-75ES Digital Audio Tape (DAT) player. The DAT player was connected to the computer using a Stealth DAI2400 digital audio interface. The interface allowed input and output between the computer and DAT player.

The echo model was approximated by a digital system. An example of how a digital system is constructed is given in section 2.5. A digital system can be realized on a computer using a simple program.

An echo model was simulated using a program which was designed to synthesize echo times for a sound at 30° as defined by the coordinate system in Figure 1.7. When the echo model was listened to, it did not stimulate localization. The system noticeably altered the perceived sound with some high frequency gain and mid-frequency attenuation, but these alterations had no effect on localization.

2.3 Closer Examination of the Echo Model

Why doesn't the echo model work?

Let us compare a real ear with the ear model proposed by Batteau. The frequency response of a real ear can be measured using a probe microphone inserted into the ear canal of a listener. An impulse sound is generated 30° from in front of the

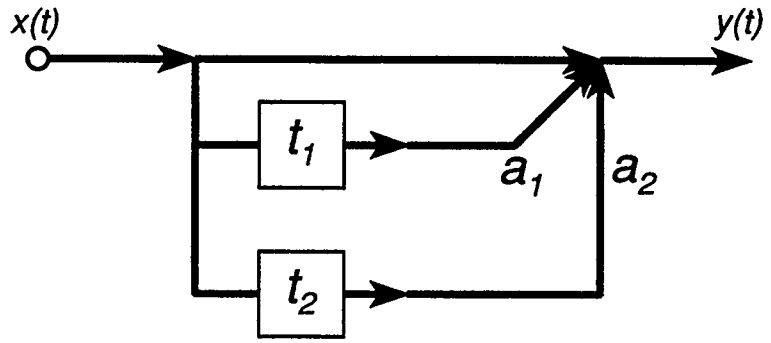

Figure 2.4: The signal flow graph for a echo model of the pinna.

listener and the impulse response of the ear is measured. Blauert's book [Bla83] lists measured ear responses on page 294. The measured response should be similar to the response of Batteau's model.

Batteau's model will be compared to a real ear in the frequency domain. Thus, the frequency response of Batteau's system must be calculated. His system is shown in Figure 2.4. The response of the system is

$$y(t) = x(t) + a_1 x(t - t_1) + a_2 x(t - t_2), \tag{2.1}$$

where $x(t)$ is the input, and $y(t)$ is the output. The parameters t_1 and t_2 are the delay times for reflectors one and two in Figure 2.3. Batteau gives these values as

$$\begin{aligned} t_1 &= 60\mu sec, \\ t_2 &= 200\mu sec. \end{aligned}$$

The parameters a_1 and a_2 are the efficiency of reflectors one and two. They determine how well the surfaces reflect sound. If their values are zero, no sound is reflected. If their values are one, the sound is completely reflected. It is assumed that the efficiencies are

$$a_1 = a_2 = 0.6.$$

The exact value of these parameters is not critical. They will not effect the overall shape of the frequency response, only the magnitude of the response. In other words, the size of a peak or dip in the response is effected, but not the location of the peak or dip. Blauert's frequency-band theory shows that the location of the peaks and dips is more important.

The frequency response of Batteau's echo system can be shown to be

$$\frac{Y(f)}{X(f)} = 1 + a_1 e^{-j2\pi f t_1} + a_2 e^{-j2\pi f t_2}, \tag{2.2}$$

where $X(f)$ and $Y(f)$ are the Fourier transforms of $x(t)$ and $y(t)$. Figure 2.5 compares the magnitude frequency response of a real ear and Batteau's system. The two do not correspond well, especially at low frequencies. Batteau's model shows a large dip at 3kHz while the real ear response gives a peak. A similar disagreement occurs at 8kHz. At the higher frequencies, things are better. The dip in the echo model at 12kHz almost corresponds with the dip in the measured response at 11kHz, and the peak and dip in the echo model at 15kHz and 18kHz almost line up with the peak and dip at 13kHz and 16kHz in the measured response. Thus, Batteau's system does not model the actual response of an ear; therefore, it does not stimulate localization.

2.4 Improving the Echo Model

Blauert [Bla83] hypothesized that the pinna is too small to reflect the larger wavelengths of sound. Recall that Batteau's data was measured on an ear model five times the size of a real ear. The echo times given in his paper were scaled down versions of the times from the large ear models. The smaller pinna of a real ear might not reflect a sound as well.

Batteau's model might be improved by assuming that only high frequencies are reflected by the ear. The large-wavelength, low-frequency components are largely

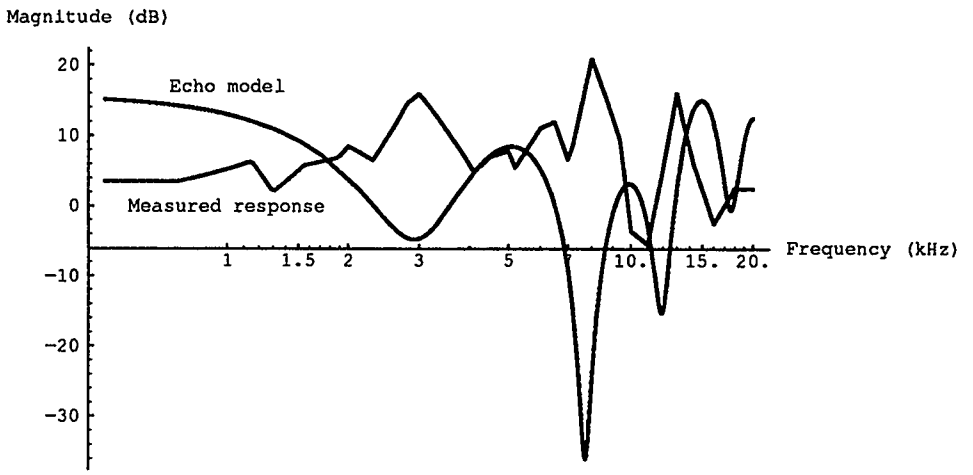

Figure 2.5: A comparison of the response from the echo model and the measured response of a real ear measured with a probe microphone.

unaffected by the pinna. Hence, the low frequencies in Batteau's model did not correspond to the measured response of an ear, where as the high frequency response corresponded much better.

2.5 Implementing an Improved Echo Model

If only the high frequencies of a sound go into the delay paths of the filter, only the high frequencies are echoed. Thus a system like that shown in Figure 2.6 should be used. It will be referred to as the *improved echo model.* The block labeled $HPF(f)$ is a high-pass filter. The improved system was realized as a digital system. The following description of how the digital system was realized will demonstrate how the first echo model was simulated.

To start, Batteau's echo model was approximated by a digital system. The sampling frequency was $f_s = 44.1$kHz. This is the sampling rate used in Compact

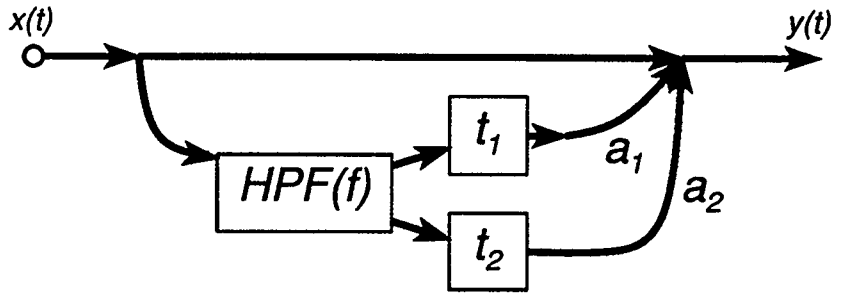

Figure 2.6: Improved Echo Model.

Disc players. Equation 2.1 becomes

$$y[n] = x[n] + \hat{a}_1 x[n - m_1] + \hat{a}_2 x[n - m_2], \tag{2.3}$$

where $x[n]$ and $y[n]$ are sampled versions of $x(t)$ and $y(t)$. The parameters $\hat{a}_1$ and $\hat{a}_2$ are finite precision values of a_1 and a_2. The values m_1 and m_2 are related to the time delays t_1 and t_2. They are the nearest integer number of sample periods to the echos times and are given by

$$m_1 = \text{Int}(t_1 f_s) \tag{2.4}$$

$$m_2 = \text{Int}(t_2 f_s), \tag{2.5}$$

where Int() is a function that rounds a fraction into the nearest integer. In the example

$$m_1 = \text{Int}(60\mu sec \times 44100Hz) = \text{Int}(2.646) = 3$$

and

$$m_2 = \text{Int}(200\mu sec \times 44100Hz) = \text{Int}(8.820) = 9.$$

The new delays are

$$\hat{t}_1 = m_1 \div f_s = 3 \div 44100 = 68.03\mu sec$$

and

$$\hat{t}_2 = 204.1\mu sec.$$

2.5.1 A High-Pass Filter

The high-pass filter passed high frequencies and attenuated low frequencies so that only the high frequencies were passed through the echo loops. A first order filter was used.

In order to design a first order high-pass filter in the digital domain, a low-pass filter was designed and then transformed into a high-pass filter. A first order low-pass filter is given by the equation

$$y[n] = 0.5(x[n] + x[n-1]). \tag{2.6}$$

This is called an averaging filter. The z-transform of this function is

$$Y(z) = 0.5X(z)(1 + z^{-1}), \tag{2.7}$$

where $Y(z)$ and $X(z)$ are the z-transforms of $x[n]$ and $y[n]$. Thus the transfer function is

$$\frac{Y(z)}{X(z)} = 0.5(1 + z^{-1}). \tag{2.8}$$

To change this low-pass system into a high-pass one, a transformation given by Oppenheim and Schafer [OS89] was used:

$$z^{-1} \rightarrow \frac{z^{-1} + \alpha}{1 + \alpha z^{-1}}, \tag{2.9}$$

$$\alpha = -\frac{\cos(0.5(\theta_p + \omega_p))}{\cos(0.5(\theta_p - \omega_p))}, \tag{2.10}$$

θ_p = low-pass filter cutoff frequency,

ω_p = desired cutoff frequency.

Applying this transformation to the averaging filter results in

$$\begin{aligned} \frac{Y(z)}{X(z)} &= 0.5\left(1 - \frac{z^{-1} + \alpha}{1 + \alpha z^{-1}}\right) \\ &= 0.5\left(\frac{1 + \alpha z^{-1} - z^{-1} + \alpha}{1 + \alpha z^{-1}}\right) \end{aligned}$$

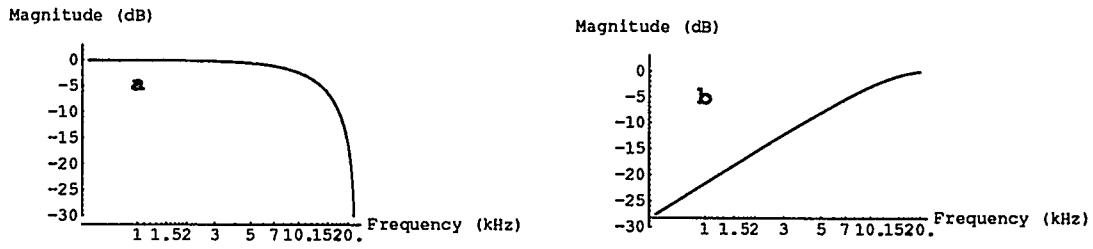

Figure 2.7: a) Low-pass averaging filter response, b) High-pass filter response.

$$= 0.5\left(\frac{(1-\alpha)+(\alpha-1)z^{-1}}{1+\alpha z^{-1}}\right)$$
$$= \frac{1-\alpha}{2}\frac{1-z^{-1}}{1+\alpha z^{-1}}. \tag{2.11}$$

The cutoff frequency is defined as the point on the frequency magnitude plot where the output is attenuated by -3dB from the peak value; thus, the cutoff frequency for the low-pass filter can be found by solving

$$-3\text{dB} = \left|0.5\left(1+e^{-j\theta_p}\right)\right|, \tag{2.12}$$
$$\theta_p = 1.568. \tag{2.13}$$

The desired cutoff frequency for the high-pass filter is 10kHz. This value has to be converted to radians. The radian value (ω_p) is related to the Hertz value (f_p = 10kHz) by the relation

$$\frac{\omega_p}{2\pi} = \frac{f_p}{f_s}. \tag{2.14}$$

Solving for ω_p

$$\omega_p = 2\pi\frac{10000Hz}{f_s} = 1.425. \tag{2.15}$$

Substituting these values into equation 2.10, the value $\alpha = -0.07441$ is obtained. This completes the design of the high-pass filter. The magnitude frequency response of the high-pass filter is shown in Figure 2.7 along with the low-pass filter response.

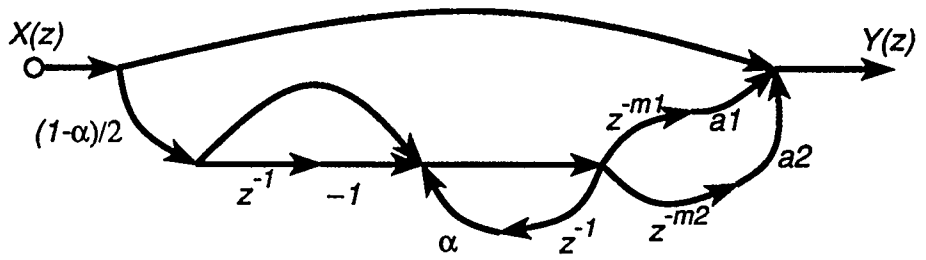

Figure 2.8: Flow diagram for new echo model.

2.6 Results Using the Improved Echo Model

Now that the high-pass filter was designed, the system in Figure 2.6 could be realized. The system in Figure 2.8 was realized by adding the high-pass filter for the $HPF(f)$ block in Figure 2.6. The transfer function of the improved echo model is

$$\frac{Y(z)}{X(z)} = 1 + \frac{1-\alpha}{2}\frac{1-z^{-1}}{1+\alpha z^{-1}}(\hat{a}_1 z^{-m_1} + \hat{a}_2 z^{-m_2}). \tag{2.16}$$

The frequency response was determined by setting $z^{-1} = e^{-j2\pi f/f_s}$ and is plotted together with the measured ear response in Figure 2.9. The low frequency response corresponds better to the measured response, but the system still does not match the actual pinna response, particularly at 8kHz where there was a large spike.

The improved echo model was realized on a computer and listened to. It sounded more pleasant than the first echo model, but no localization was stimulated. The notch at 8kHz was perceptible and made the sound seem unnatural.

2.7 Towards a Better Model

Neither of the echo models stimulated sound localization in the listener, but some progress was made by trying to improve the delay model. Batteau suggests that the ear can be more accurately modeled by a continuous echo model. Such a model

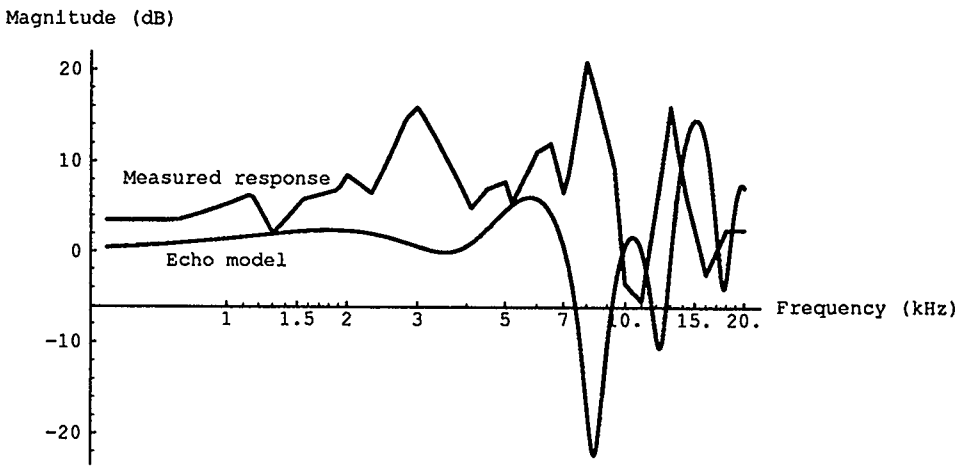

Figure 2.9: Response of the new echo model compared with a measured pinna response.

would have the equation

$$\int_0^{T_{max}} a(t)e^{-st}dt = \int_0^{\infty} a(t)e^{-st}dt = A(s). \tag{2.17}$$

As Blauert [Bla83] notes, this equation is just the Laplace Transform. Any LTI system can be represented with such an echo model, but it would require an infinite number of delay elements. Thus, it is impossible to realize.

The echo model could be improved by more carefully modeling the physics of the pinna. The improved echo model was derived in this way. It was noticed that the ear was too small to reflect large wavelengths effectively and so the low frequencies were filtered out. By looking at similar effects, like diffusion and resonance, the pinna model might eventually stimulate localization. This is a difficult fluid dynamics problem, and is best solved using careful measurements on real ears. This is discussed in the next chapter.

Chapter 3

Head Related Transfer Functions

A better model of how the external ear alters a sound was required since an echo model did not stimulate localization. A better method of solving this fluid dynamics problem is to measure the response of a human listener's ear. When sounds are introduced to the ear, accurate measurements of the pinna distortions can be made. This was briefly described in the previous chapter, but will be dealt with in more detail here.

3.1 Head Related Transfer Function Measurement

There is a standard method of measuring ear responses, and it is best described by Blauert [Bla83]. A subject is seated in an anechoic chamber, which is an echo free room. The subject's head is restrained so head movement will not be a factor (this is also for the safety of the subject, because a fine tipped probe microphone is inserted into their ear. The tip of the probe sits at the entrance to the ear canal and head movement could damage the ear drum). Sounds, usually white noise or impulse functions, are produced at specific angles around the listener. The frequency response of the pinna system is measured using the microphone in the subject's ear. The subject is then removed from the room and the measurements are repeated. The difference in the two measurements is the transfer function of the pinna. A transfer function measured in this way is called a Head-Related Transfer Function (HRTF).

It is important to remember that the pinna, and thus the HRTF, of every individual is different. Researchers will average the response of several subject's HRTFs to try and determine an "ideal" pinna response. Fisher and Freedman [FF68] and

Butler and Belendiuk [BB77] show that people can localize while listening through other peoples pinna, and thus it is assumed that a pair of "ideal" ears exist that give good localization cues to everyone. Butler and Belendiuk confirm such a concept by showing that some people can localize better when listening to sounds recorded through other peoples pinna. However, the averaging of pinna responses is not always effective. Blauert [Bla83] discusses how the peaks and dips in the magnitude response, features that are important to localization, can be blurred by mathematical averaging. These peaks and dips occur in all pinna responses, but occur at slightly different frequencies. In mathematical averaging, the sharpness of these features would be destroyed. A more sophisticated means of generalizing the responses is required.

Kendall, Martins and Decker [MD89] visually averaged many pinna responses to obtain transfer functions for their research. This is a better method. However, they did not reproduce their results in their paper. The transfer functions chosen for this thesis are found in Blauert's book [Bla83] on page 90–91. The responses from 25 subjects are averaged together using complex averaging. This means that the real and imaginary components of the response were averaged separately.

3.2 Linearity, Causality and Time-Invariance of HRTFs

In order for HRTFs to have any meaning, the pinna system has to be causal, time-invariant, and linear. Is the human ear such a system?

If it is assumed that the ear is a rigid body with constant dimensions and composition, then physics dictates that the system is linear, causal, and time-invariant. But is the ear a constant rigid body? In section 1.4 it was seen that the ear adjusts to the volume of the sound presented to it. It will dampen and respond to frequencies differently depending on the intensity of the sound. But this does not

make the ear non-linear or time-variant. The damping and frequency variations are caused by the cochlea, and the cochlea does not effect the measurement of the pinna's characteristics.

An argument against the causality of the ear is possible. Sometimes people get "ringing ears." When no sound is present the ear will start to hear a ringing sound. But this does not mean the HRTFs are non-causal. Again, the ringing effect is caused by the cochlea and the HRTF measures the pinna response.

A new question arises out of this discussion. Can localization be stimulated by duplicating the effects of the pinna alone. This is the question contemplated in the remainder of the chapter.

3.3 HRTF Notation

Since several HRTFs were used, a notation was developed to describe them. A HRTF will transform a sound from one location to another. It therefore needs a *reference point* and a *destination.* The function will "move" a sound from the reference point to the destination, and in this thesis sounds will be moved in limited ways. They will be moved around in the horizontal plane, and from inside the head to outside the head. A sufficient notation is

$$\begin{aligned} HRTF &= H_{\text{reference,destination}}(f), \quad \text{where} \\ H &= \text{the name of the function} \\ f &= \text{frequency variable} \\ \text{reference} &= \text{angle of reference (0–360°, } h\text{, or } e \text{)} \\ \text{destination} &= \text{angle of destination (0–360°, } h\text{, or } e \text{).} \end{aligned} \tag{3.1}$$

If the value of the reference or destination angle is listed as h, this means In Head Localization (IHL) is perceived. If the value of the reference or destination is listed

as e, this means the location is at the entrance to the ear canal. If the reference or destination is an angle, the location is at that angle. For example, a transfer function that will move a sound from inside the head to outside the head and to the right of the listener could be called $A_{h,90}(f)$.

3.4 HRTFs and Localization Cues

HRTFs will be used to take into account all the cues that are presented to actual ears. The duplex cues will be represented in the measurements because the phase will contain the Interaural Time Difference (ITD) information and the magnitude response will contain the Interaural Intensity Difference (IID) information. The pinna information will of course be present since this is what was measured. In addition, the effects of other possible cues, such as shoulder reflections and torso reflections, will be accounted for. Exact duplication of the input to the ear drum that occurred when natural localization took place is the goal. Since the only audio input to the brain is from the ear drum, it was assumed that localization would be duplicated. If an artificially reproduced sound is identical to a naturally occurring sound, it is called an *ear-equivalent* sound. Blauert [Bla83] believes that if HRTFs are reproduced in the playback of sound, proper localization will result because the sound is ear-equivalent.

Wightman and Kistler [WK89a][WK89b] showed HRTFs can stimulate localization. They put subjects in an anechoic chamber and measured their HRTFs as described above. The ability of the listener to locate sounds in a free-field environment was measured. Then they measured the same ability using headphones and the HRTFs. The headphone sounds were made in the following fashion. First, for every individual in the experiment, the HRTF for each position was measured ($H_{e,\theta}(f)$). Then the HRTF for headphones to the ear drum was measured ($H_{e,h}(f)$). The

response could be simulated by the overall transfer function

$$H_{h,\theta}(f) = \frac{H_{e,\theta}(f)}{H_{e,h}(f)}. \tag{3.2}$$

Wightman and Kistler found that localization in the headphone-simulated environment corresponded extremely well with free-field localization, except that the number of front-back reversals increased. Thus they proved HRTFs could stimulate sound localization.

There is one problem with Wightman and Kistler's research in terms of this thesis's objective. In their experiment, the HRTFs were measured for every individual. This would be impractical for common use. The time and equipment needed to measure an ear response can not be easily made available to the general public. It would be preferable to have one ideal HRTF that can be used by any listener.

Another problem is that Wightman and Kistler ensured that every listener had ten hours of practice locating sounds in a free-field environment. The average listener will not have the time or means to train themselves to locate sounds. So the effect of HRTFs on untrained listeners must be investigated. Can untrained people localize sounds encoded with ideal HRTFs?

3.5 Testing HRTF Localization

Blauert [BL78] constructed a system that fulfills this thesis's first objective. He makes a system to simulate stereo, free-field speaker listening on headphones. He does this by using a system of HRTFs which were not measured for each individual. He believed the system made listening to head phones more pleasant and natural; more like a free-field speaker environment. Thus he proved that ideal HRTFs can stimulate localization.

The system that he described in his paper was duplicated, and sounds were processed with it. Localization was definitely obtained. The sound image moved

forward, away from the ears. The most noticeable effect was on sounds that appeared directly in the ear. One annoying result of the sound stage of headphones is that sound can appear right at your ear. This is similar to having a mosquito try to enter your ear. After processing with Blauert's system, this annoyance disappeared.

A sound processed with Blauert's system was played for six untrained people in an informal manner. First the original sound was listened to, then the processed sound. The listeners had difficulty telling the difference between the unprocessed and processed sound; except when the sound was right in their ear. Most listeners said the sound moved away from their ears but could not tell where it went. They did not perceive a sharply defined sound stage in front of them, and although the sound moved away from the ears, it did not consistently move outside the head.

Since it was noticed that the listeners had trouble comparing the processed and unprocessed sound, the presentation was changed. The original presentation played the entire unprocessed sound and then the entire processed sound. By the time the listeners heard the processed sound, they had forgotten what the original had sounded like. It was concluded that the human brain, unless trained to do so, can't retain a memory about the features of a sound for very long; at least the features that were important for localization. To compensate for this lack of memory, the HRTFs were switched on and off every five seconds, and the listeners would hear the difference between the processed and unprocessed sound during each transition. When the subjects heard this presentation, they instantly heard the difference between the sounds, and all the subjects noticed the sound stage was moved away from the ears. A better sound stage was reported, but not a sharply defined sound stage.

HRTFs have been proven to stimulate localization; however, the sound stage is not sharply defined and is not outside the head when perceived by untrained listeners. It is proposed that the HRTFs in Blauert's paper [BL78] were poorly

measured. Better instruments for measuring HRTFs are available now than were available at the time his paper was written. By improving the HRTFs it might be possible to improve the sound stage.

3.6 Free-Field Listening System Using New HRTFs

Blauert's book [Bla83] contains HRTFs measured with more modern equipment. These will be used to build another free-field simulator for headphones. To build a free-field system, digital filters that duplicate the HRTFs $H_{0,30}(f)$ and $H_{0,330}(f)$ were implemented. The magnitude and phase of the HRTFs and the response of the digital filters is shown in Figures 3.1 and 3.2. Finite Impulse Response (FIR) filters with 250-taps were used.

FIR filters are systems that have a finite impulse response. In other words, there exists some number $M < \infty$ such that

$$h[n] = 0 \text{ for all } n > M, \tag{3.3}$$

where $h[n]$ is the impulse response of the system. The filters are designed by determining the digital impulse response of a system from its frequency response. This is accomplished using the inverse Discrete Fourier Transform (DFT):

$$h[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{H}[k] e^{j(2\pi/N)kn}, \tag{3.4}$$

where $\hat{H}[k]$ is a sampled version of the frequency response. Since $H(e^{j\Omega})$ is periodic, only one period of it needs to be sampled. The number N is the number of sample points in the period. In this case, $N = 250$. If the system depicted in Figure 3.3 is built, then an FIR filter with the desired response is obtained. This method of designing an FIR filter is called *frequency sampling.*

$H_{0,30}(f)$ an $H_{0,330}(f)$ are the difference between a sound straight ahead of the listener and a sound at 30° or 330° from straight in front of the listener. Thus the

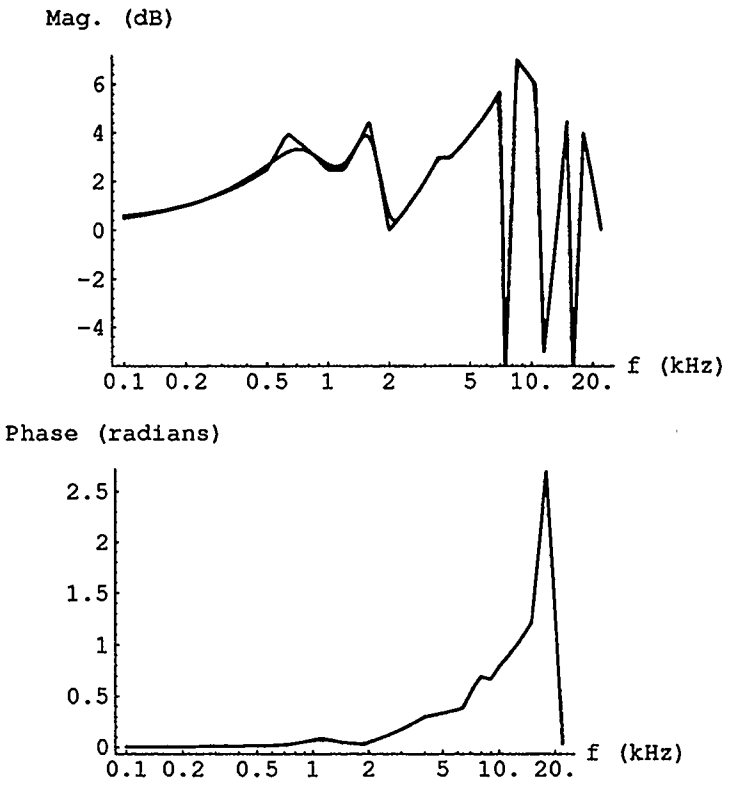

Figure 3.1: Magnitude and phase response of $H_{0,30}(f)$. Jagged response is measured HRTF and smooth response is the FIR filter.

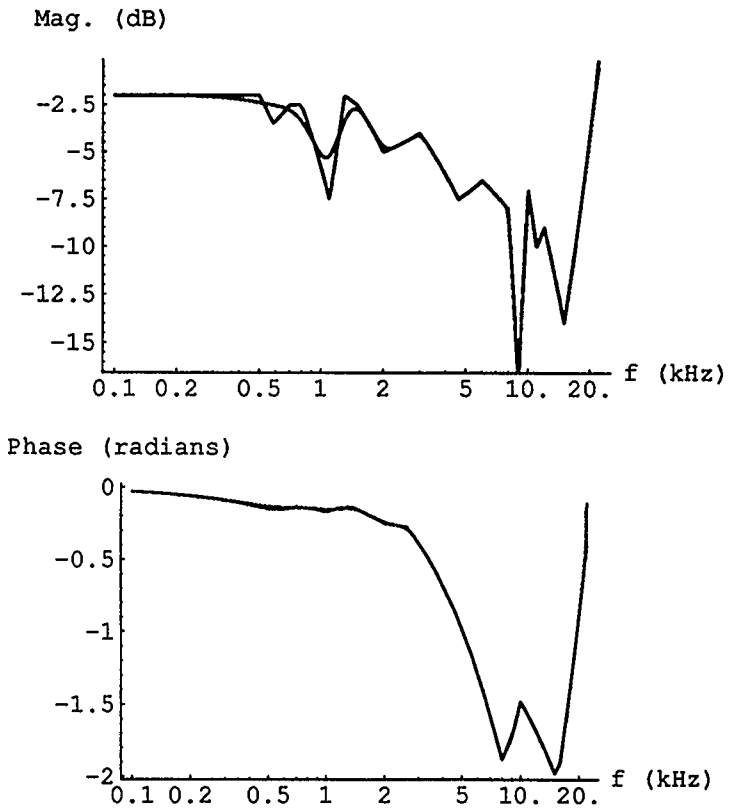

Figure 3.2: Magnitude and phase response of $H_{0,330}(f)$. Jagged response is measured HRTF and smooth response is the FIR filter.

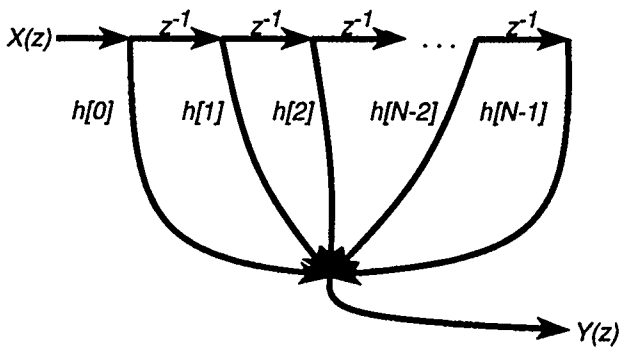

Figure 3.3: An example of a discrete-time, FIR, filtering system.

reference point for these HRTFs is directly ahead of the listener. A transfer function from a headphone's perceived location to directly ahead of the listener is required. Then the reference point can be at the normal headphone listening location. This HRTF, called $H_{h,0}(f)$, is shown in Figure 3.4.

Note that $H_{h,0}(f)$ attenuates the mid-frequencies. If a sound in front of the subject is listened to, and then the same sound is listened to on headphones, the headphones would appear to have a boosted mid-range. Blauert [BL78] confirms that headphones alter sound in this way. By using $H_{h,0}(f)$ more natural sounding mid-frequencies will be heard when using headphones.

Figure 3.5 shows the free-field simulator that was implemented. The stereo sound is input into the system and transferred from inside the head to directly in front of the listener and then from in front of the listener to either 30° or 330°. Note that the left hand signal reaches the right ear and vice versa. This is called cross-talk and is essential for free-field simulation.

The new system did not stimulate localization better than Blauert's system. The sound stage was moved away from the ears but not outside the head. The principles used in both simulators was the same, just the HRTFs used were different. Blauert's system had a high pitched notch in its sound that the new system didn't, and the new system sounded more muddy than Blauert's; as if the bass was boosted too much. This is an example of the large variance in the HRTFs available in the literature. The variance between the HRTFs used in Blauert's system and those used in the new system was quite audible, but the new HRTFs did not improve localization.

3.7 Multi-Channel Localization Cues Using HRTFs

Blauert [BL78] points out that the same method used in section 3.6 can be used to make a multi-channel system by adding more HRTFs. In section 3.6 sounds were

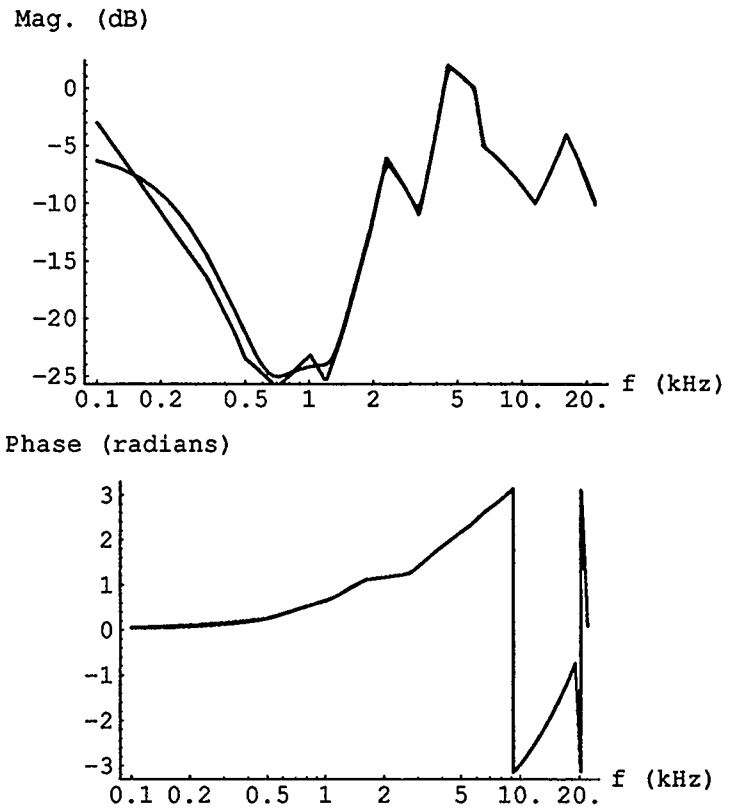

Figure 3.4: Magnitude and phase response of $H_{h,0}(f)$. Jagged response is measured HRTF and smooth response is the FIR filter.

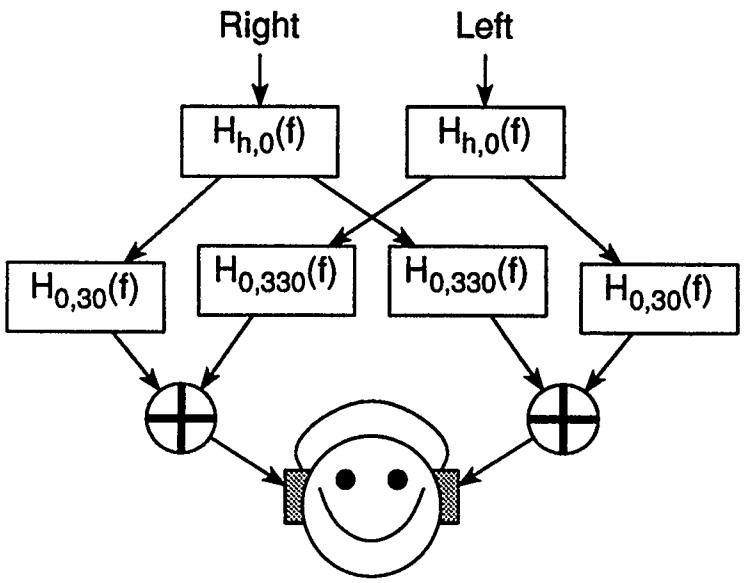

Figure 3.5: System for simulating free-field listening on headphones.

located at 30° and 330°. In this section sounds will be located at two additional positions: 150° and 210°.

3.7.1 Testing a Multi-Channel System

A multiple-subject test was designed to test HRTFs that locate sounds both in front and behind of the subject. The test involved playing sounds at one of six locations: 30°, 150°, 210°, 330°, and directly at the left and right ears. This is depicted in Figure 3.6. Six random noise signals were used. The system for free-field listening in section 3.6 was used to transform each of the six sounds to 30°, 150°, 210°, and 330°. $H_{0,150}(f)$ and $H_{0,210}(f)$ are shown in Figures 3.7 and 3.8, and as in the previous section, digital FIR filters with 250-taps were used to duplicate the transfer functions. Sounds were also played in either the left or right speaker of the headphones. This produced in-head localization directly in either ear (ie: lateralization).

Each subject was seated in front of a computer. All six sounds were played at each location in a random order, for a total of thirty six sounds. Then the test

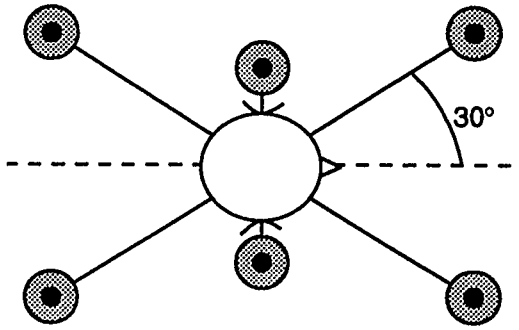

Figure 3.6: Six locations of sounds for the experiment.

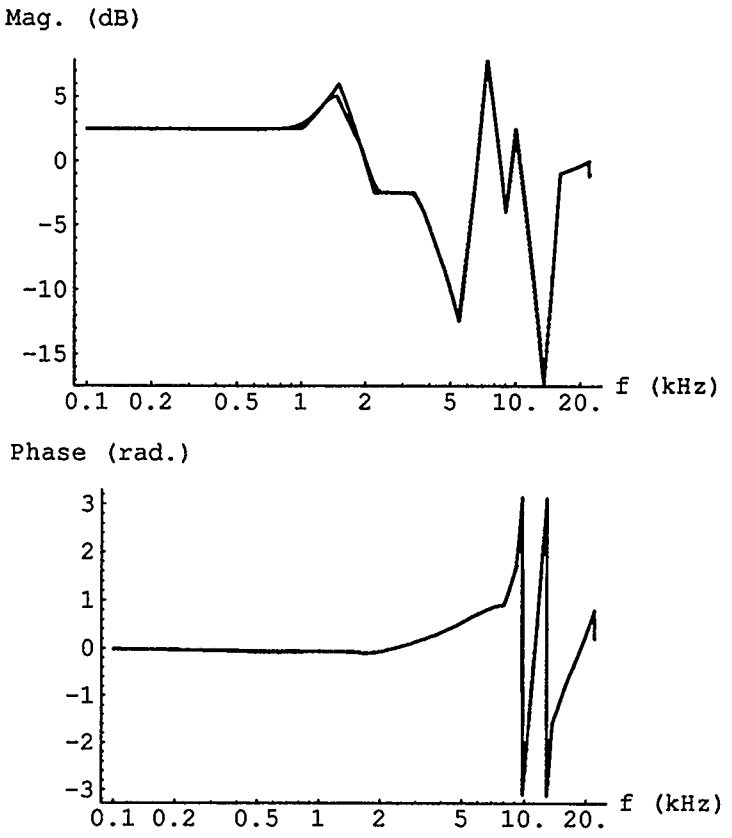

Figure 3.7: Magnitude and phase response of $H_{0,150}(f)$. Jagged response is measured HRTF and smooth response is the FIR filter.

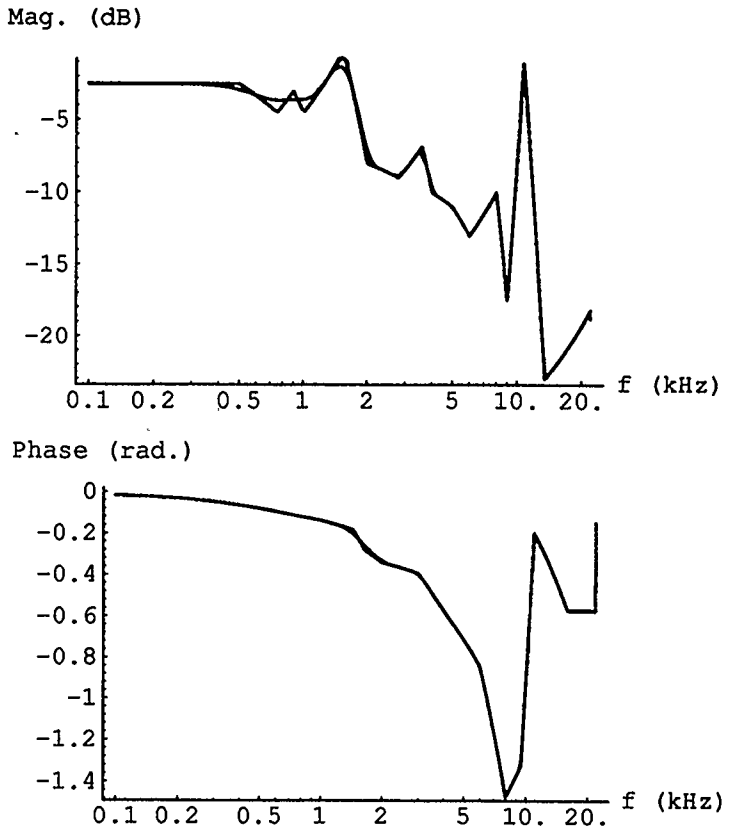

Figure 3.8: Magnitude and phase response of $H_{0,210}(f)$. Jagged response is measured HRTF and smooth response is the FIR filter.

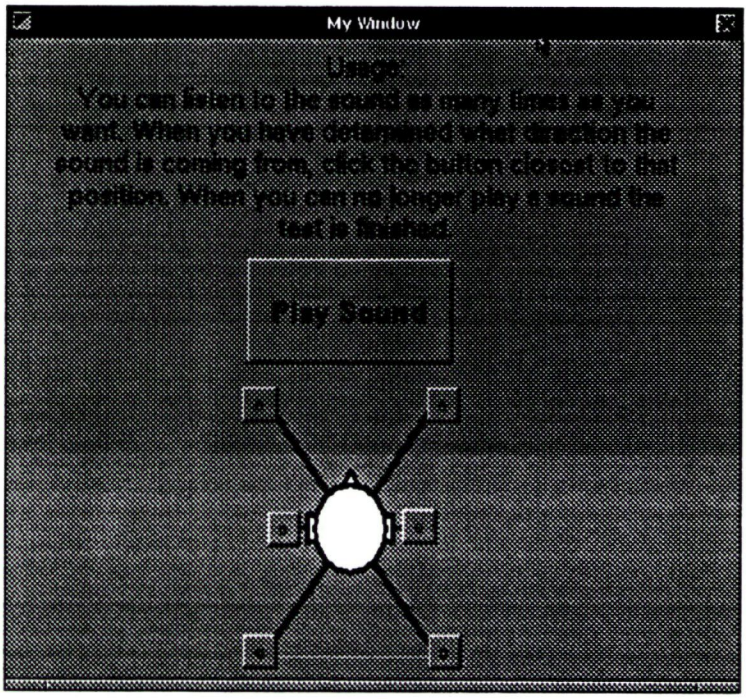

Figure 3.9: Computer display for experiment.

was immediately repeated. The subject could listen to each sound as often as they wanted, then they selected a position on the computer screen that most suited the perceived localization. The computer display is shown in Figure 3.9. The subjects were not told whether the decisions they made were right or wrong.

3.7.2 Results of Multi-Channel Test

The six possible sound locations can be analyzed as only three. The sound in the right ear and the left ear are identical. The only difference is that one is on the left side and the other is on the right. Since there were no left/right confusions in the experiment, the differences between a right and left ear presentation can be ignored. The same argument can be applied to the 30°/330° and 150°/210° positions. Thus there are only three locations possible: front, center, and rear. Each judgment was specified by two variables. The first variable specified the "correct" localization, and

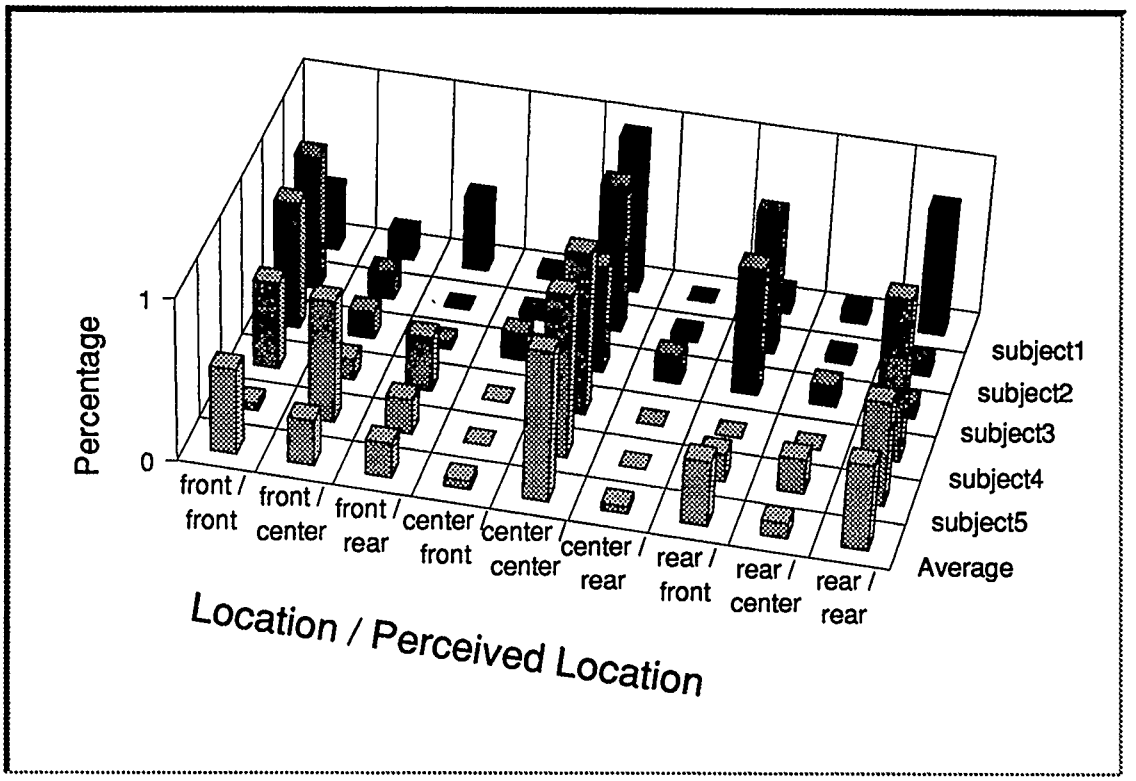

Figure 3.10: Results from experiment.

the second specified the perceived localization. As a short hand, this will be written (location)/(perceived location). This notation is used in Figure 3.10 to show the results of the experiment.

Five subjects were tested. On average, the subjects perceived the desired location. This shows that localization was taking place; however, the deviation of the results is very large. For example, subject 5 did not recognize any forward locations while subject 2 recognized them all. All the subjects were very good at recognizing the center location, but were confused about the difference between front and back signals. This is to be expected because Wightman and Kistler showed front-back reversals are common in headphone listening.

3.7.3 Effects of Learning in the Multi-Channel Experiment

The effects of learning were investigated.

Did the subjects learn to tell the difference between sounds in front of them and behind them during the experiment? It was hypothesized that if the subjects heard a "front" sound and then a "rear" sound and compared them, they would be able to determine the difference between them; especially if the proper localization cues were present. To test this hypothesis the results from the first thirty-six sounds were compared with the results from the second thirty-six sounds. No improvement was noticed; therefore, the subjects did not learn to localize during the experiment.

The subjects were never told where any sound they heard was located. Thus, they had no reference except their stored memory of naturally occurring sounds. The subjects failed to localize well and they failed to learn to localize well. It can be assumed that the sounds they were hearing did not correspond to sounds they naturally hear.

3.7.4 Conclusions From Multi-Channel Experiment

From the experiment, it was concluded that headphone localization using HRTFs is not effective for multi-channel listening. It did not provide the subjects with enough information to tell the difference between front and back locations. This is puzzling because it is in direct contradiction to Wightman and Kistler's results. Something must have differed in the way the experiments were performed. The difference should be determined so the results can be improved.

Chapter 4

A Listening Environment For Outside-the-Head Localization

Inside-the-Head localization(IHL) was the most prevalent problem in the HRTF tests. This chapter is concerned with trying to achieve Outside-the-Head Localization (OHL).

Two papers provide insight into what causes OHL. The first is the work by Sakamoto, Gotoh, and Kimura [SGK76]. It explains how reflected energy can be used to locate a sound outside the head. The second source is the work of Plenge [Ple74]. Plenge shows that knowledge of the environment is important to OHL.

4.1 Echoes and OHL

Sakamoto, Gotoh, and Kimura used a model of a human head with artificial pinna. They placed the head model in an anechoic chamber with a speaker placed in front of it. Microphones in the model's ears recorded sounds played over the speaker. This process encodes the sounds with the HRTFs of the head model's pinna. This is very similar to the experiments in Chapter 3. The only difference is that in Chapter 3 a digital filter was used instead of a head model to encode the HRTFs. When Sakamoto, Goth and Kimura played the recorded sounds to listener's wearing headphones, the sound should have been localized in front of the listener. Sakamoto, Gotoh, and Kimura found that the percentage of outside-the-head localizations was very low. Thus they observed the same phenomenon that was observed in Chapter 3; HRTFs do not promote OHL.

Sakamoto, Gotoh and Kimura suggest the problem is that the recordings were made in an anechoic chamber. Sound are located inside the listener's head because

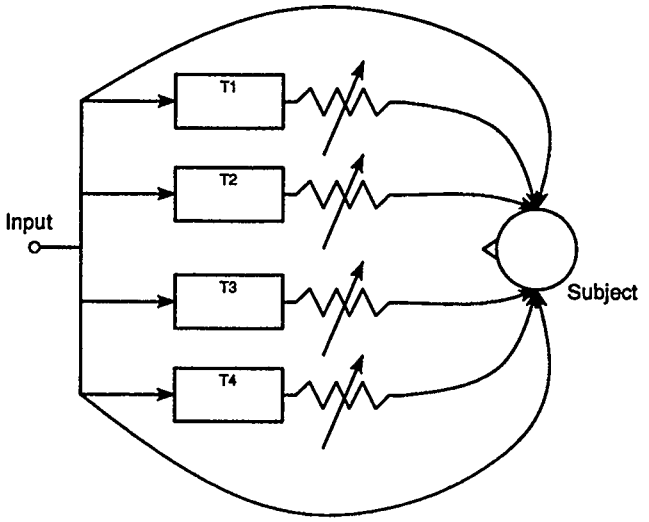

Figure 4.1: Echo system for OHL.

an anechoic environment is not natural. The system explored in the previous chapter suffers from the same problem because the HRTFs used were measured in an anechoic chamber. To get the sound out of the head Sakamoto, Gotoh, and Kimura added echoes to the signal to simulate the effect of sounds reflecting off walls. The percentage of OHLs perceived by the subjects increased.

The echo system from Sakamoto, Gotoh, and Kimura's paper was duplicated. It is depicted in Figure 4.1, where the value $T1, T2, T3$, and $T4$ are delay times. Each delay time was adjustable to values between 5msec and 20msec, and the gain of each delay path was also adjustable. OHL was achieved with this system, but the echoes made sounds tinny; as if listening through a cardboard tube. Since the echoes were causing a the tinny sound, some other way of inducing OHL had to be found. Sakamoto, Gotoh, and Kimura suggest that OHL is stimulated by reflected energy in the signal, and the echoing effect they used is only one way of introducing reflected energy into a signal.

4.2 Reverberation and OHL

Plenge [Ple74] found that signals produced over headphones can be located inside the head or outside the head. Signals played over headphones that are located outside the head are called *ear-adequate.* Those that are not are called *non-ear-adequate.*

In an experiment, Plenge recorded a sound using a normal microphone in an anechoic chamber. The sound was the speech of someone familiar to a group of subjects. When the subjects used headphones to listen to the sound they perceived the auditory event inside their heads. Thus a signal recorded in an anechoic chamber seems to be non-ear-adequate. Note that in his experiment no HRTF information was encoded in the sounds. It is not clear that HRTFs recorded in an anechoic environment are non-ear-adequate from this result, but Sakamoto, Gotoh, and Kimura showed this in the previous section. Plenge's first experiment merely strengthens Sakamoto, Gotoh and Kimura's observation.

The second experiment performed by Plenge was more interesting. He recorded familiar speech in a normal (not anechoic) room using a model head with microphones in the ears. The sound was played back using headphones and the subjects located the sound outside the head at the location of the original source. Thus, Plenge succeeded in localizing a sound outside the listener's head at a specific point in space. What did Plenge do that wasn't done in Chapter 3? He added two things to the sound. The HRTF information was added by the model head, but as Sakamoto, Gotoh, and Kimura showed, this is not sufficient for OHL. The second element added was reverberation. Reverberation is the echo-like effect the walls of a room impose on a sound. It adds reflected energy to the sound. Thus Plenge's results show that reverberation is necessary for OHL.

The findings of Plenge, Sakamoto, Gotoh, and Kimura are complementary. They both point out that reflected energy is required for ear-adequate sounds. Plenge

used reverberation while Sakamoto, Gotoh and Kimura used a tinny-sounding echo system. Reverberation is based on the acoustics of rooms and therefore sounds more natural. Plenge's second experiment suggests that adding the reverberation will improve localization.

4.3 A Localization System With Reverberation

Plenge used a concert hall to measure parameters for a localization system. Recordings of sounds played over speakers were made in the hall. The recordings were made using model heads to add pinna information, and were analyzed to determine the HRTFs and find the level and timing of the hall's reverberation parameters. Then artificial HRTF cues and artificial reverberation were added to a sound. The processed sound was played on headphones to listeners who were given a drawing of the concert hall. Plenge does not state whether or not the listeners had ever visited the hall. The subjects all located the processed sound outside the head. The exact position of the localization was not very accurate, 60% of the localizations being within a $\pm 15°$ arc of the actual location of the sound.

This test proves reverberation contributes to OHL, but Plenge's system did not resolve the location of the sound very well. As seen earlier, the angular blur in the auditory space is of the order of a few degrees. A 30° blur is far too large. The HRTFs used in Plenge's system were very crude. They were not measured with the accuracy available using todays technology. Using better HRTF measurements, better localization should be obtained.

Reverberation was added to the HRTF free-field listening system from section 3.6. The new system is shown in Figure 4.2. The only difference is the addition of the reverberation algorithm which is described in section 4.4. Twenty people listened to the new system. The sound listened to was the song *Sgt. Pepper's Lonely Hearts*

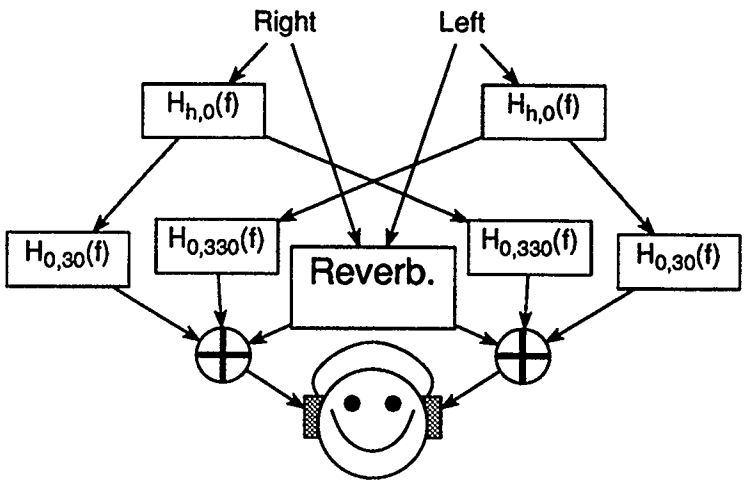

Figure 4.2: Free field listening system with reverberation.

Club Band by the Beatles. This is a fairly old recording which brings out the flaws in headphone listening. The effect was switched on and off every 5 seconds during the presentation so that the listener could tell the difference between the two sounds. The listeners noticed that the sound was improved. They noted that the sound moved outside the head and made listening much more pleasant. But only half the people could actually locate the new sound stage as being in front of them. The rest located the sound as "somewhere out there," so it is suggested that improving the HRTFs using reverberation did not improve the localization blur.

Though the sound was not sharply located, the system still meets the requirements of our first objective. A method of free-field listening on headphones has been obtained. But there is still the question of simulating multi-channel environments on headphones. A localization system using HRTFs and reverberation has too large of an auditory blur to realize a multi-channel system.

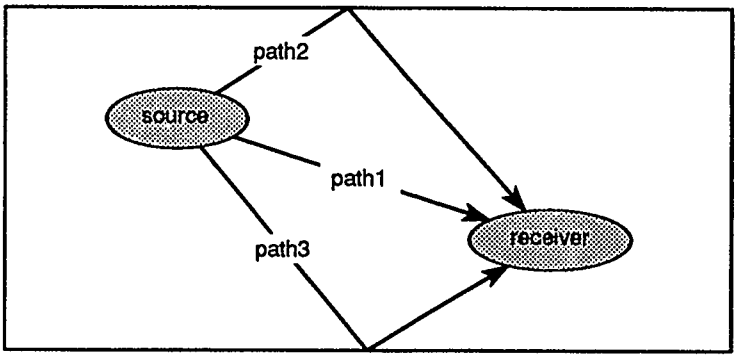

Figure 4.3: The many sound paths in a room.

4.4 Reverberation

Reverberation is the effect the environment has on a sound; particularly when the environment is an enclosed room. When an object produces a sound in a room, the sound will reach the receiver through many paths as shown in Figure 4.3. Note that all the paths are of different length. Sound traveling along each path will reach the listener at different times. The composite of all these sound paths is the reverberation of the room. If the surfaces that are reflecting the sound in the room do not absorb any sound (ie: are perfect reflectors) then the sound will reverberate for a very long time. But every real room has some sound absorbing material, so reverberation will stop after a fixed amount of time. The more sound absorbing material in a room compared with its total volume, the sooner the reverberation will stop. The reverberated sound will decay exponentially and the time it takes for the sound to die out is called the *reverberation time* of the room. This is an important reverberation parameter. A large concert hall can have a reverberation time of over three seconds. A room in an average house (about $10m^3$ in volume and average sound absorption) has a reverberation time of under one second.

4.4.1 Reverberation by Convolution

An algorithm to simulate reverberation using digital technology is required. The most obvious way to accomplish this is to measure the reverberation of an actual room and use it to create artificial reverberation. This method comes directly from digital signal processing theory. If the room is a system, it is characterized by measuring its impulse response.

The reverberation in a room can be modeled with a linear, causal, time-invariant system. It is easy to show that a reverberation system meets all these requirements. First, the system must be linear because reverberation is a linear combination of the original signal. It is assumed that the room doesn't change over time (not significantly) so the system is time-independent. No sound will be heard in the room unless a sound source initiates it; therefore, the system is causal. Another interesting property is that all real rooms have a finite reverberation time; thus, the system must have a finite impulse response.

A reverberation system can be realized by convolving a sampled sound with the impulse response of a room. If the room's impulse response is $h[n]$ and the sampled sound is $x[n]$, Equation 1.17 describes the output $y[n]$. This approach is computationally intensive and not easily realizable with todays technology. Fortunately, more efficient algorithms have been devised.

4.4.2 Schroeder's Reverberation Algorithm

The fundamental work on digital reverberation was done by Schroeder [Sch62]. He was one of the first researchers to make a digital reverberation system. The main goal in designing a reverberation system is to get a good echo density while having good tone. Echo density is described as the number of reflections occurring in one second. A typical value in a room is about 1000 reflections per second.

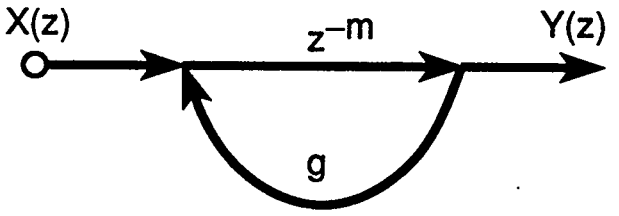

Figure 4.4: Feedback comb filter.

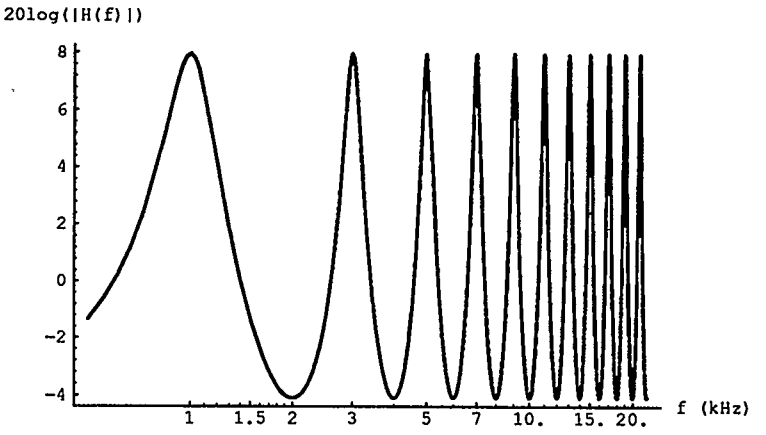

Figure 4.5: Feedback Comb filter magnitude response: $g = 0.6$, $m = 22$, $f_s = 44100$Hz.

Schroeder's first attempt at building a reflection unit was a feedback comb filter as shown in Figure 4.4. Sound is fed back and added to the original after being delayed. Then this accumulated sound is fed back and added again and again. The coefficient $-1 < g < 1$ dampens the sound over time; simulating the absorption of sound. But this filter has poor tone. The frequency response of such a filter, shown in Figure 4.5, has many spikes. These spikes do not sound pleasant.

The response of the filter can be flattened by modifying it to be the filter in Figure 4.6. This is an allpass filter and its frequency response is shown in Figure 4.7. The response is flat; no colourization of the sound takes place.

Feedback filters are still used by Schroeder because they produce echoes necessary

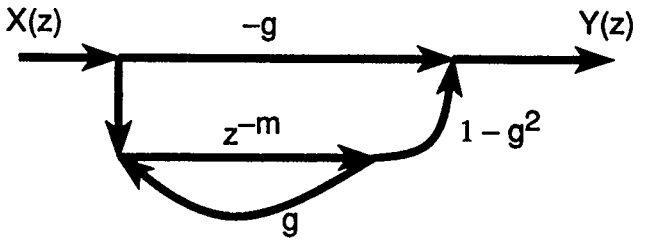

Figure 4.6: Allpass filter.

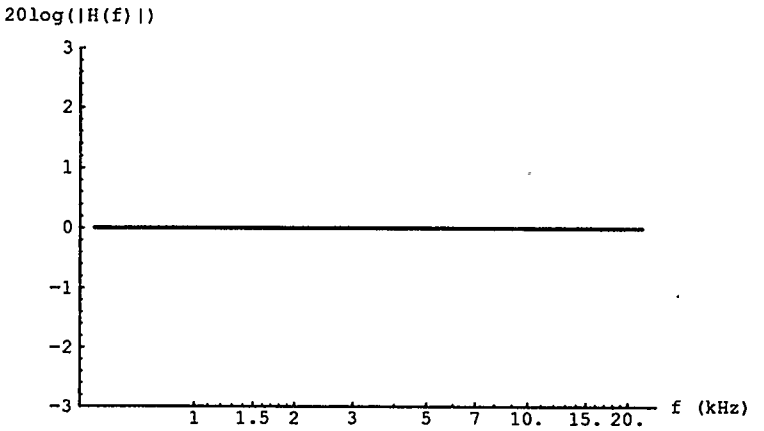

Figure 4.7: Allpass filter magnitude response: $g = 0.6$, $m = 22$, $f_s = 44100$Hz.

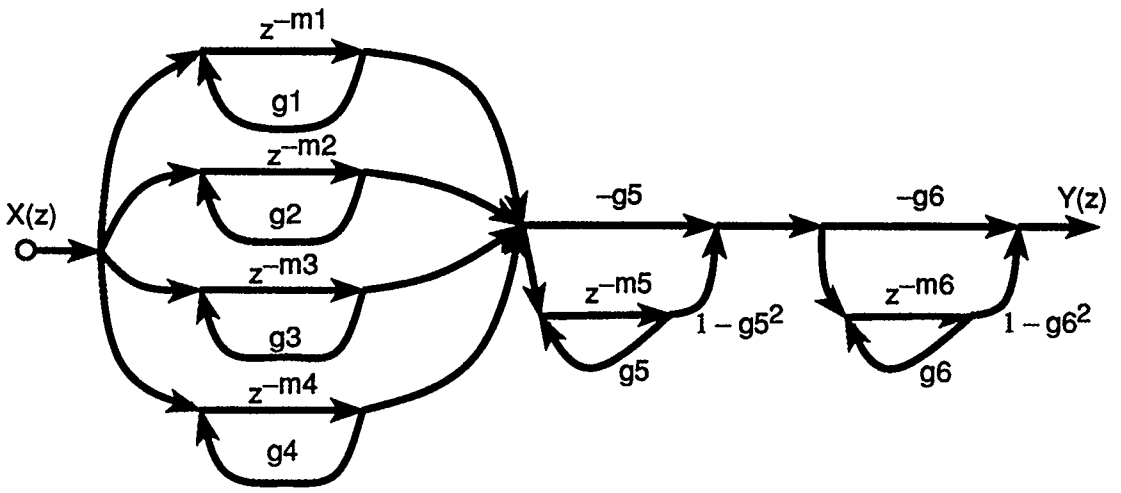

Figure 4.8: Schroeder's reverb system.

for good reverberation. By using several feedback filters in parallel, the magnitude response evens out and the tone is improved. But it is wasteful to use only feedback filters. The delay time in a feedback filter is about 40ms. This only produces 25 reflections in a second. Thus 40 parallel feedback filters are required to produce an echo density of 1000. Connecting them in series produces a terrible sound because the colourization of the tone is magnified. To increase the echo density, allpass filters are used. Schroeder uses four parallel feedback filters in series with two allpass filters to produce the desired echo density. The system is shown in Figure 4.8.

To make the system stereo, Schroeder suggests mixing the inputs together and mixing the outputs of the feedback filters in different phases. Noting that there is no difference between having the allpass filters before or after the feedback filters (because the system is linear), a stereo reverberation can be realized as in Figure 4.9.

Schroeder's reverberation system sounded much better than the direct echo system used by Sakamoto, Gotoh, and Kimura, but it did not sound natural because it still had poor tone.

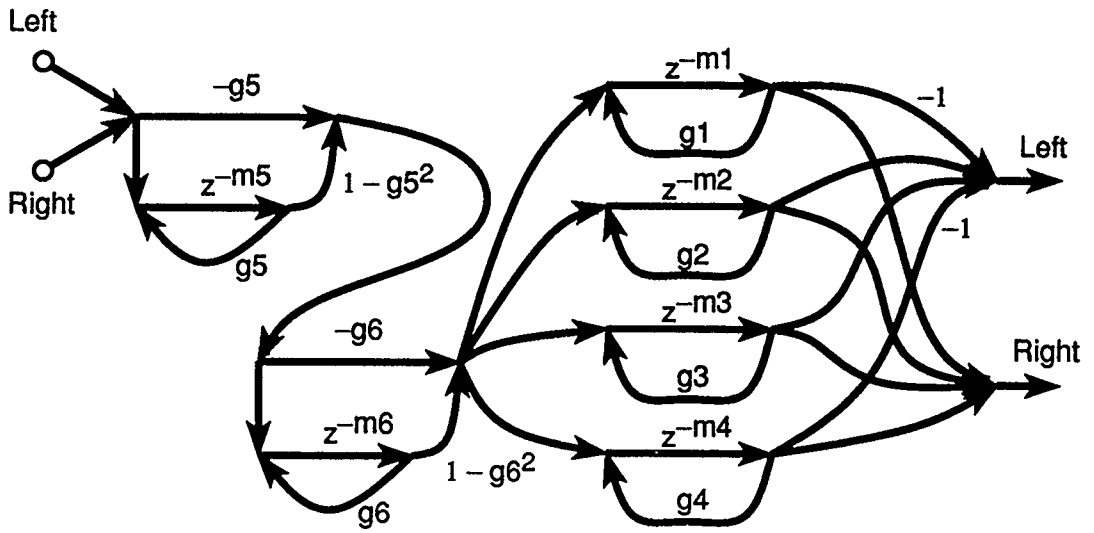

Figure 4.9: Schroeder's reverb system modified for stereo.

4.4.3 Moorer's Reverberation Algorithm

Moorer [Moo79] noted this same problem with Schroeder's system. He developed a different system by modifying the feedback filter. It is modified by putting a low pass filter in the loop. The resulting filter is shown in Figure 4.10. The design of this filter is very similar to the design of the improved echo model in Chapter 2. Moorer uses six of these filters in parallel, and one allpass filter in series. The overall system is shown in Figure 4.11. This system can be modified for stereo sound in the same way Schroeder's was modified. Moorer's system was realized in software and sounded much better than Schroeder's. It is the algorithm used for the free-field headphone system in this chapter.

4.5 Realization of the Free-Field Simulator in Hardware

The free-field simulator was realized in a real time system because the computer simulation of the algorithm was time consuming. First, the desired sound had to be sampled into the computer. Then the sound had to be processed with a program.

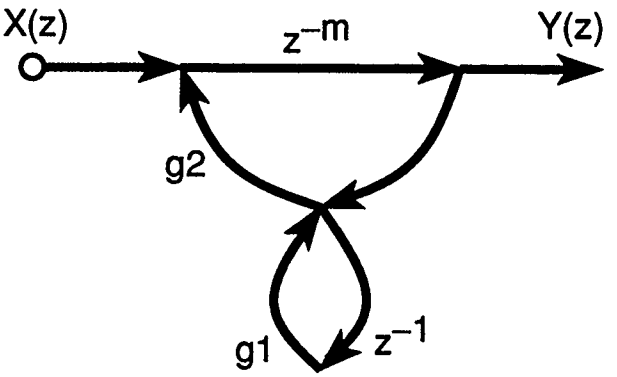

Figure 4.10: Modified comb filter.

Then the sound had to be transferred from the computer back to the recorded medium. So a device that performs an equivalent function without a noticeable pause to the listener is desired. Such a system is called a *real time* system.

This section will briefly describe some of the design decisions and practices used in implementing the free-field simulator in hardware. A real time system uses dedicated hardware devices to perform its operations. In this thesis, Xilinx 4005/10 programmable logic devices [Xil91] were used. A significant effort was made to simplify the simulator algorithm so it could be implemented on these devices.

The hardware system is composed of three elements as shown in Figure 4.12. An interface board controls input from a Compact Disc (CD) player and output to a Digital Audio Tape (DAT) player. The interface communicates information to an FIR device implemented on two Xilinx 4005 chips. The FIR chip then passes information onto a reverberation device implemented on a 4010 chip. In the figure, all signal paths represent stereo audio signals.

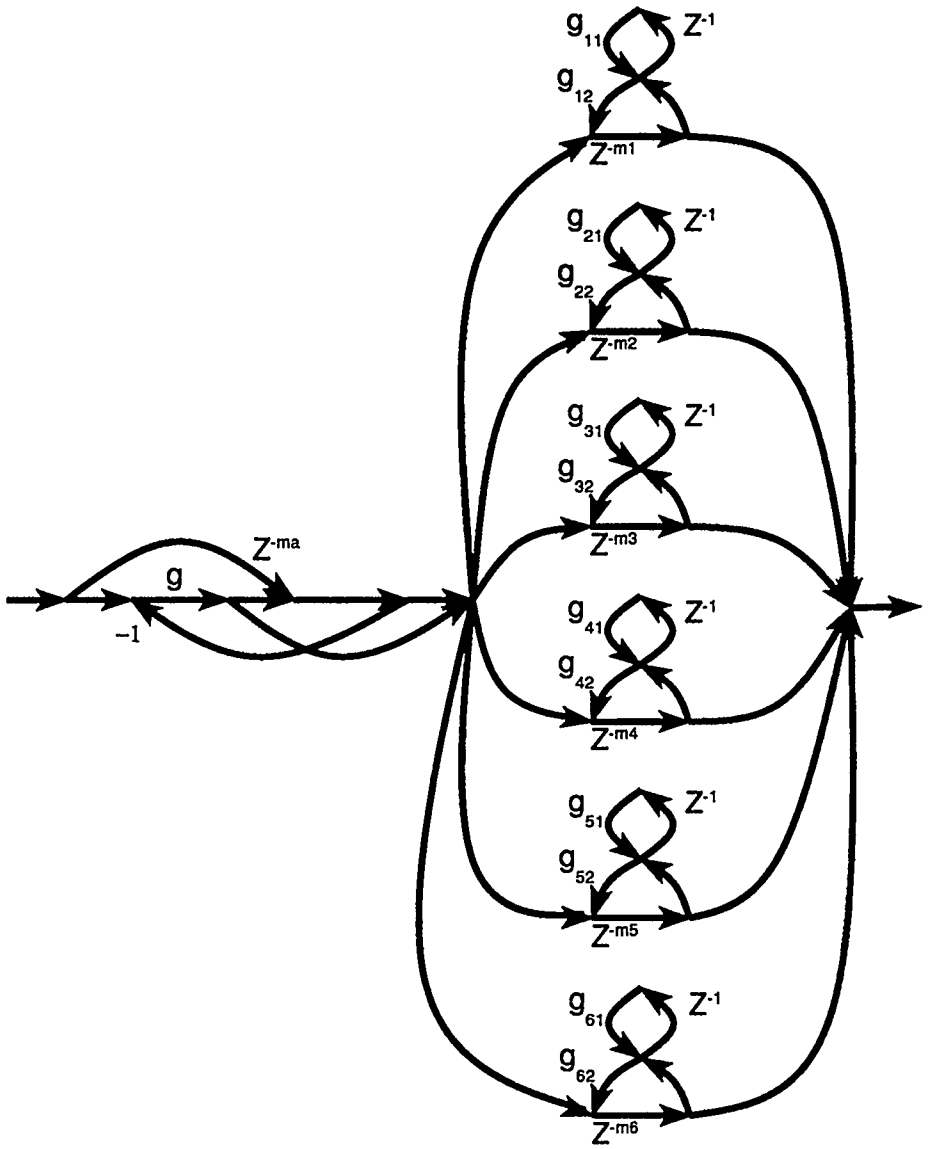

Figure 4.11: Moorer's reverberation system.

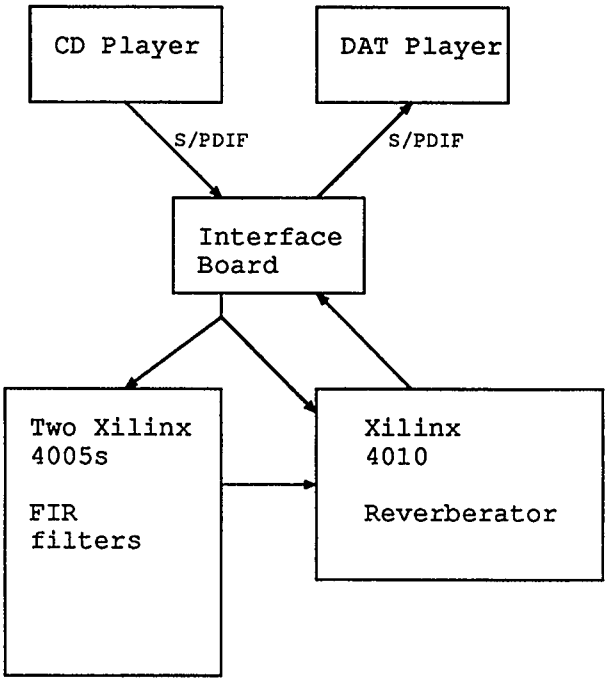

Figure 4.12: Overview of the hardware system.

4.5.1 Interface Board

The signal output from the CD player, and input to the DAT player, is in the Sony/Phillips Digital Interface Format (S/PDIF) [Cry92]. It is a serial signal with encoded clock information. The interface board recovers the music data and the clock from the S/PDIF signal. It also performs the inverse function; given music data and a clock signal it creates an S/PDIF signal.

Crystal Semiconductors's 8402/8412 chip set [Cry92] was used to design the board. The 8412 receives S/PDIF signals and the 8402 transmits S/PDIF signals.

4.5.2 FIR Design

The Finite Impulse Response (FIR) filters had to be realized in an efficient manner. The Xilinx chip contains a limited amount of resources; too limited for a direct implementation of the 256-tap FIR filters used in the computer simulations. Thus

coefficient	$H_{h,30°}(f)$	$H_{h,330°}(f)$	*coefficient*	$H_{h,30°}(f)$	$H_{h,330°}(f)$
h[0]	4	-6	h[16]	-3	7
h[1]	4	-11	h[17]	-20	23
h[2]	-1	-5	h[18]	-2	0
h[3]	3	3	h[19]	1	-6
h[4]	-2	5	h[20]	-14	-5
h[5]	-5	0	h[21]	15	-18
h[6]	9	1	h[22]	6	7
h[7]	6	3	h[23]	15	-10
h[8]	-4	-1	h[24]	-14	22
h[9]	-5	-3	h[25]	-7	-2
h[10]	-2	-5	h[26]	-4	9
h[11]	-7	1	h[27]	6	-12
h[12]	6	-7	h[28]	6	-6
h[13]	30	-29	h[29]	0	-11
h[14]	12	-3	h[30]	0	-5
h[15]	-11	4	h[31]	4	0

Table 4.1: Coefficients for FIR filters.

a more efficient FIR algorithm was designed by limiting the number of taps and the size of the multiplication coefficients.

The program NOMAD [KGST92] can design FIR filters with fewer taps and smaller coefficients. The frequency sampling method of FIR design is very inefficient by comparison. NOMAD uses an adaptive algorithm called simulated annealing to find a set of FIR coefficients. The desired HRTF responses ($H_{h,30°}(f)$ and $H_{h,330°}(f)$) were input into NOMAD, and it was found that both filters could be realized using 32, 6-bit coefficients. The coefficients are listed in Table 4.1. The coefficients have been scaled to facilitate NOMAD design. The coefficients for $H_{h,30°}(f)$ have been divided by ≈ 1.85877, and the coefficients for $H_{h,330°}(f)$ have been divided by ≈ 0.56234.

These two 32-tap FIR filters are still very large compared to the 4005 resources. To maximize the use of resources and make the design small enough to fit on two 4005s, a bit-serial design was chosen. In a parallel design, a number is represented

as a collection of bits that simultaneously appear on parallel lines. Sixteen bits on sixteen parallel lines would be required to represent a number. In a bit-serial design, one line is used to represent a signal and the sixteen bits are transmitted sequentially on the line. This is one example of how bit-serial design helps to reduce resource usage.

To further reduce the size of the design, no multiplications were used. Multiplications require a great deal of resources. Since the coefficients contain very few bits, it is more efficient to shift and add values. For example, shifting a value up by one bit is the same as multiplying the value by two. Thus, the FIR filter design contained only shift and add elements. These elements are efficiently implemented in bit serial designs.

The FIR filters were designed to process stereo signals. The left and right channel inputs were multiplexed in time. In other words, first a left sample was input, and then a right sample. By altering the FIR filter in Figure 3.3 so that $z^{-1} \rightarrow z^{-2}$, a stereo filter was produced. The free-field simulator is realized by combining the two FIR filters as in Figure 4.13. Note that the outputs of the two filters are added together only after delays are applied. This is so that the right sample's $H_{h,30^\circ}(f)$ is added to the left sample's $H_{h,330^\circ}(f)$ and vice versa. The multiplexer makes sure that the proper additions are sent to the output.

The system was designed using the DFIRST [Gra89] bit-serial design system. Once the DFIRST system was designed and tested, it was translated into a Xilinx input file by a program called TRANS [GT92].

4.5.3 Reverberation Design

The reverberation system was described in Figure 4.11. The values for the various parameters are shown in Table 4.2. To implement this system it was necessary to

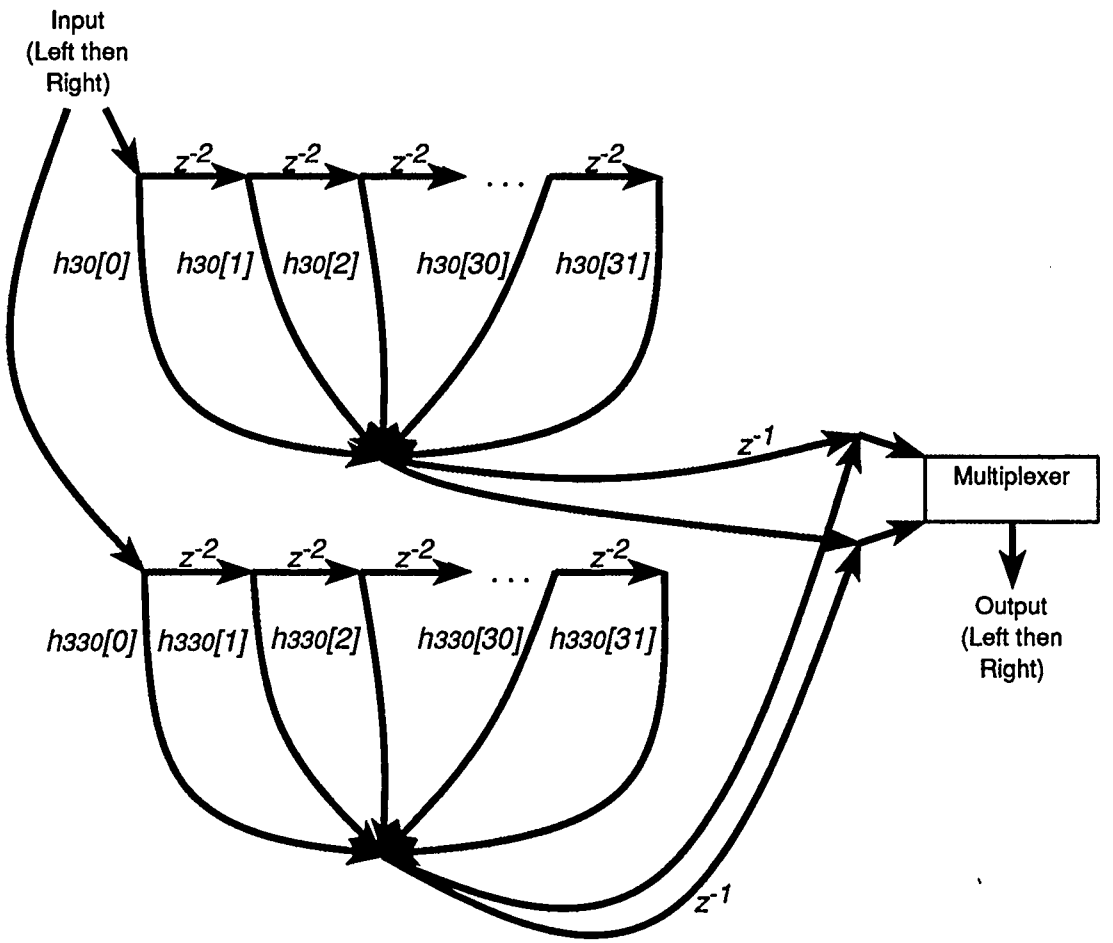

Figure 4.13: Realization of FIR system for free-field localizer.

parameter	value
$g_{11}, g_{21}, g_{31}, g_{41}, g_{51}, g_{61}$	0.25
$g_{12}, g_{22}, g_{32}, g_{42}, g_{52}, g_{62}$	0.5
g	0.5
$m1$	2048
$m2$	2304
$m3$	2560
$m4$	2816
$m5$	3072
$m6$	3328
ma	256

Table 4.2: Parameters for reverberation.

interface Random Access Memory (RAM) chips to the Xilinx chip. The reverberation algorithm requires the storage of more values than the Xilinx chip would allow. So memory external to the Xilinx chip was added. The reverberation chip also contained a summation network as shown in Figure 4.14. It allowed the listener to make the reverberation louder or quieter.

The reverberator was designed in two stages. The first stage was a DFIRST design. DFIRST was used to make the bit-serial design and all the filtering elements. Then the system was translated into a LOGSIM [KTBW89] design using TRANS. LOGSIM is a gate-level logic simulator, and the RAM interface was added at the gate level. Then the finished LOGSIM design was translated into a Xilinx input format using TRANS again.

4.6 The Free-Field Simulator System

The FIR and reverberation algorithms used in the hardware system are different than the ones used earlier in the chapter. To insure that they would still stimulate localization, the altered algorithms were implemented on a computer. Localization

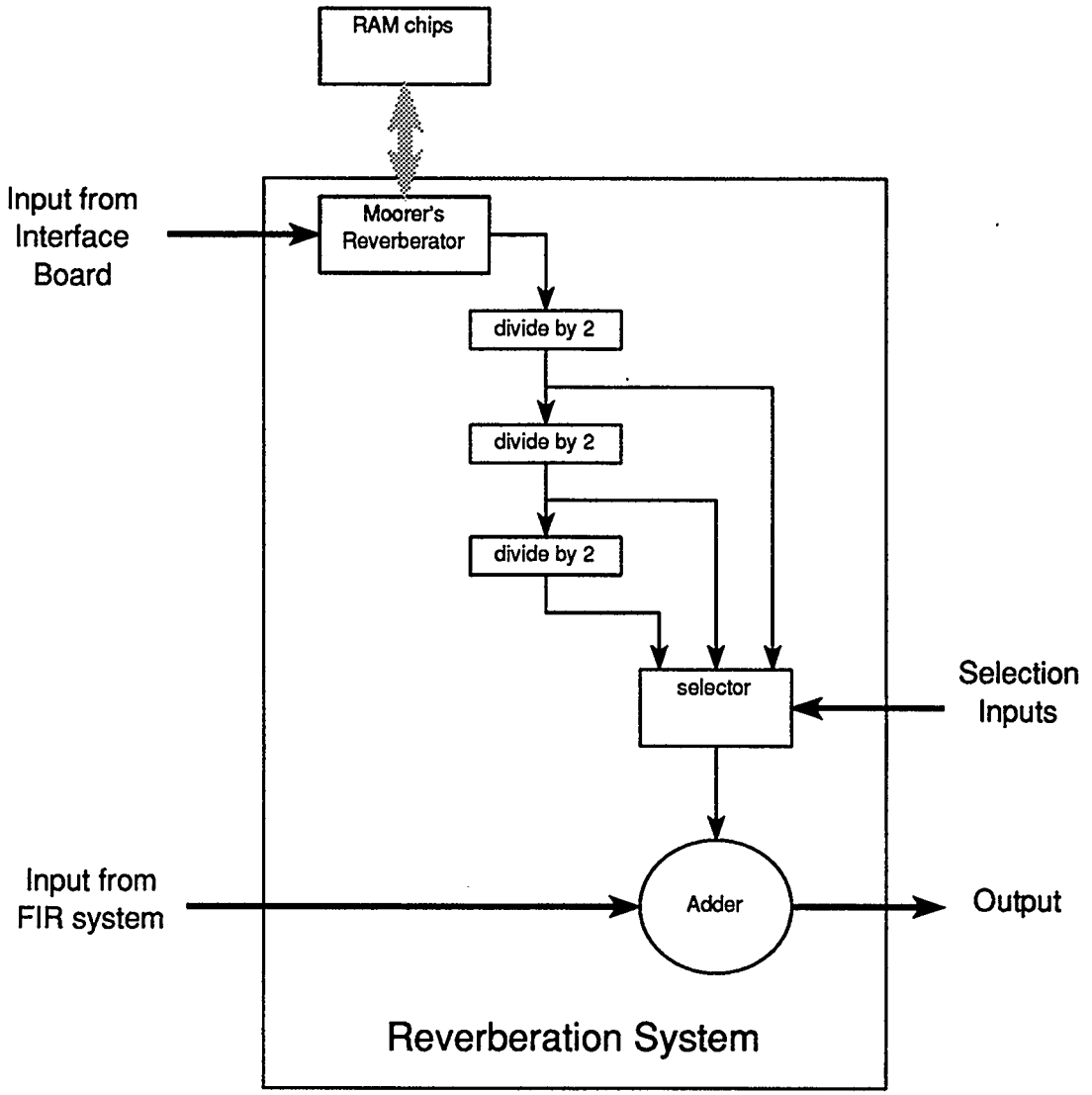

Figure 4.14: Summation network in the reverberation system.

was not compromised by the new algorithms, so realization in hardware commenced. The system was assembled and behaved the same as its simulation.

Chapter 5

Defining an Environment

When HRTFs were first used, it was assumed the auditory system could be modeled as in Figure 5.1. This means the ear is the only input to the hearing system, and any auditory event can be duplicated by duplicating the ear input. But it was found that this was not the case. There are many inputs to the interpretive center of the brain.

5.1 A New Model of Sound Localization

It is proposed that a better model of the interpretive center is shown in Figure 5.2. In this model the effects of learning and memory are included. A distinction between memory and learning is made because one is short term and one is long term. A person learns a sound, or environment, while they are listening to it. If they memorize this sound so that it can referred to later, then they do not have to relearn it because it is in their memory.

It is hypothesized that the new model better predicts localization and will show how a multi-channel recording system on headphones can be realized.

5.1.1 A Contradiction

A contradiction exists in the experimental results of the previous chapters. The experiments of Wightman and Kistler [WK89a][WK89b] showed that people could

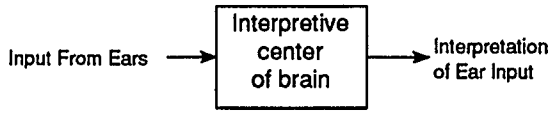

Figure 5.1: Flow diagram of hearing system.

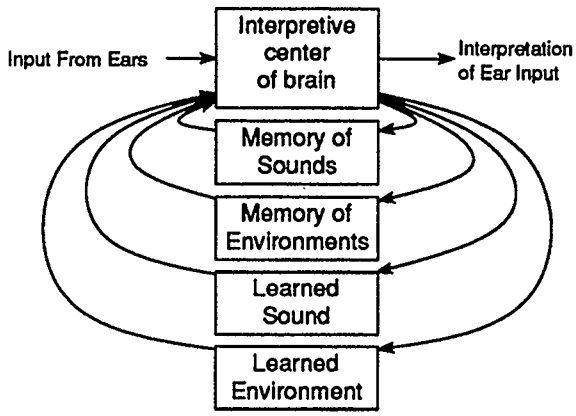

Figure 5.2: Improved flow diagram of hearing system.

locate sounds recorded in an anechoic chamber and presented on headphones. The sounds were perceived to be located outside the listeners head. But Sakamoto, Gotoh, and Kimura [SGK76] showed that people could not locate sounds recorded and presented in this manner. The sounds were located inside the listeners head.

One might suspect the difference in the results is caused by the pinna responses. Wightman and Kistler measured the response of every individual in the experiment. This meant that the pinna responses were very accurate and tailored to the individual. Sakamoto, Gotoh, and Kimura measured the response on a general model head; therefore, the pinna response was not tailored to the experimental subjects. This could mean that sounds have to be presented from the actual ears of the individual to get good results, and only poor results could be obtained with model heads.

Blauert [Bla83] shows that model head presentation has had many problems since it was introduced. The main problem was front-back confusion, which was the problem in Chapter 3. Blauert suggests that the problem lies in the method of averaging the pinna responses to get an optimal pinna response. But Fisher and Freedman [FF68] and Butler and Belendiuk [BB77] observed that people can listen through the ears of others. Butler and Belendiuk also showed that some people can

localize better when listening to sounds recorded through other peoples pinna. Thus, some other discrepancy must exist.

5.1.2 Adaptation to the Environment

How can people listen through other peoples pinna? Carr [Car66] reports that if the ears are altered, the individual will adapt to the new shape of the ear and localize just as efficiently. In some interesting experiments, tubes and horns of brass were inserted into subjects' ears and affixed to the head. The subjects wore the headgear for about a week and then reported back for testing. It was found that after the learning period, the individuals could locate just as efficiently. This is very similar to Fisher and Freedman's experiments; except the time period is shorter.

Fisher and Freedman inserted tubes into the ears of the subjects with artificial pinna on the ends of them. The subjects did one test where they moved their heads and one where they held them still. They adapted to the new set of pinna and then they could locate with them.

But what does any of this have to do with the experimentation of Sakamoto, Gotoh, Kimura, Wightman, and Kistler? *Before performing their experiment, Wightman and Kistler made sure that their subjects had at least ten hours of practice localizing sounds in the anechoic chamber.* In Sakamoto, Gotoh, and Kimura's experiment, the listeners were never in the anechoic chamber; much less given a chance to practice localizing in one. Thus it is proposed that Wightman and Kistler's subjects localized well because they were familiar with the environment and pinna they were listening through. The subjects in Sakamoto, Gotoh, and Kimura's experiment were not used to the environment or the pinna used, and didn't localize well. To improve localization, it is hypothesized that the subjects must have an opportunity to adapt to the simulated environment.

5.1.3 Ear-Adequate Signals Revisited

Plenge's [Ple74] first experiment with head models showed that adaptation is necessary. He recorded sounds using a model head in a room and then had the subject listen to the sound *in the same room.* The listener could adapt to the environment because they were physically located in the environment. The resulting localizations were outside their heads and accurately placed; even though the actual pinna of the individual were not used. But Plenge's second experiment seems to contradict the proposal. Sound was recorded in a concert hall and then the listeners, who had not been in the concert hall, located the sound outside their heads.

An explanation is simple. A concert hall is an environment that most listeners can relate to. People have a clear idea of what a concert hall sounds like. It is something within their stored experience. The subjects were even told that the recordings were made in a concert hall and were given a picture of it. Thus they were able to localize the sounds outside the head. An anechoic chamber is a very unnatural environment. It is beyond the experience of most people. So in Sakamoto, Gotoh, and Kimura's experiment the sound was located inside the head.

Plenge confirms this hypothesis in the his paper. He states that:

> Lateralization or verged-cranial localization [in head localization] is likely to occur if (1) the short-term storage does not contain information that can be applied (missing or deficient knowledge of sound sources and sound field), and/or if (2) the signals (stimuli) are such that they cannot be related to any of the stimulus patterns stored in the long term storage.

Short-term storage refers to learned environments and sounds. Long term storage is the memory of sounds and environments. Thus it has been shown that the model of Figure 5.2 is a valid model of the hearing system.

Ericson and Agnew [EA90] noted that prior experience in listening to HRTF encoded sounds plays a key role in being able to localize. Experienced listeners learned to resolve front/back reversals through repeated exposure to the stimulus. If a listener can adapt, they can resolve the most difficult of localization problems. The subjects have to be able to learn and adapt to the sound environment they are in or have the environment stored in their memory. If they can not adapt to it and don't already know it, they can not locate in it.

A cautionary note should be made. All sound localization is not learned. There are certain innate localization abilities that all people share from birth. Peiper [Pei63] showed a sharp click will make a baby respond by turning towards it, even when the baby is only a few minutes old. Thus the learned localizing abilities are refinements to the existing system; however, these refinements are required to make a sound localize well.

5.2 Adaptation and Localization

Our new model explains why the tests performed in chapter 3 were not successful. In the first experiment, which shall be called experiment *A*, HRTFs were used to localize sounds. The HRTFs were recorded in an anechoic chamber; hence, they were non-ear-adequate. Most people can't understand an anechoic environment because it is not in their stored experience. The second test in chapter 4 added reverberation to the sounds to make them more ear-adequate. This shall be called test *B*. The test was more successful than the first because it achieved OHL. But the test did not provide good localization because the sound stage was not well defined. According to the new model, the sound stage was not well defined because the environment was not learned. If the subjects were given an opportunity to learn the environment, then localization should have improved. The new model suggests a new test, called test

C, in which the subject will be allowed to learn about a multi-channel environment. This should provide better localization results.

But how can the subjects adapt to the environment? In a real environment, people hear sounds and then compare their location with the location from other senses. For instance, they walk into a room and hear a ringing noise. They also see someone with a bell. Then they localize the sound to be where the person with the bell is. In a few seconds they have gathered enough information to understand the sound environment. Lots of information was already stored in their mind, like what a bell sounds like. This type of information makes it easier to adapt to new surroundings. But adding visual clues to localization is beyond the scope of this thesis. Not having visual clues will hinder the ability of people to locate, but there should be a way to locate without them. Moore [Moo82] describes blind people locating objects very well; therefore, audio cues presented in a defined environment should be able to allow a person (of full senses) to locate objects.

5.2.1 Head Movement and Sound Movement

Turning the head is a good way to get information about a sound. By turning their head, a listener gets a sound profile. They know the effect of their pinna on a sound for a given direction and they know what directions they are sweeping through; thus, they can learn about the sound and the environment. A system that would track the subject's head movements and adjust the sound accordingly would allow the user to learn about, and adapt to, the environment simulated. The previously mentioned work of Loomis, Hebert, and Cincinelli [LHC90] proves these cues are sufficient.

What if instead of moving the head, the sound was moved. Instead of the head moving in space, the sound would move in space. One problem is that the subject would not have feedback about the direction. When a person moves their head, they

know which way they are facing and how much of an arc they have swept through. When a sound moves around someone's head, it is not possible to get this feedback. Thus less information is provided by a moving sound.

If the movement is very simple and the subject is told how the sound is going to move, it may be possible for them to get enough information to learn about the sound and environment. One of the simplest movements is a sound circling around the head of the listener at a constant speed. This is easy to describe to the subject and easy for them to understand.

The speed of revolution of the sound is important. Aschoff [Asc63] performed an experiment in which sound was revolved around the head using speakers. If the sound rotated too fast, it would only fade from side to side. If increased further, the sound had no direction. Hence, the sound can not revolve too fast or too slow As was mentioned earlier, the brain can not remember the characteristics of a sound for a great length of time. In chapter 3, when subjects listened to a song before and after processing, they had difficulties telling the difference between the two. If the sound moves too slow, the listener will not be able to compare the sound at various positions efficiently. So in the experiment, the sound was rotated at different speeds so the effects of rotation speed could be observed.

5.2.2 Good and Bad Localizers

The concept of good and bad localizers, discussed by Wenzel, Wightman, and Foster [EMWF88], shows that some subjects consistently localize better than others. The data of experiment *A* shows this trend. The phenomenon can be explained in terms of the new hearing model (Figure 5.2). Some people do not have as much memory of sounds and do not adapt as well. The reasons are not clear. It is proposed that seeing is a far more dominant sense than hearing. Hence, some people

do not develop their hearing skills and do not localize sound well.

5.2.3 Impact of Visual Information On Auditory Localization

Moore [Moo82] shows how conflicting visual and audio localization cues are always resolved in favor of the visual cues. The hearing system will actually change its auditory localizations so that they match the visual cues. For example, if a subject saw someone moving their mouth as if they were talking and heard the actual voice through a speaker 15° off to one side, the sound eventually would be located at the moving mouth. Then if the conditioned subject closed their eyes, they would localize all sounds a few degrees off to one side. The subjects entire auditory environment is recalibrated to match the visual information. This shows how visual cues are more powerful than audio cues. It also shows how people adapt to their environment.

5.3 An Experiment Using Adaptation to Improve Localization

The results of test C should be able to confirm that localization improves if the subject can learn. Rotating sounds will be used to teach the subjects about their environment. But what judgement should the subject be asked to make. The main decision in test A was whether the sound was in front, behind, or in the center. The center location was only for control purposes and can be eliminated from test C because no one had trouble localizing it in test A. The subject should be able to resolve front-back confusions, a known problem for HRTF presentation and a problem in test A. As was noted earlier, head movement can resolve front-back reversals. It is proposed that sound movement will help eliminate front-back confusion in a similar fashion.

It is undesirable to let the subjects know what is being tested. If they know they are supposed to resolve sounds in front and behind them, it might effect their

responses. So they will be asked to say whether the sound is rotating around their head clockwise or counterclockwise. If they can't tell the difference between a sound in front of them and behind them then they will only hear a sound moving from left to right. But if they can resolve the difference between a sound in front of them and behind them, they will be able to determine if a sound is rotating clockwise or counterclockwise around their head.

The test was set up as follows. Thirty-six sounds were sampled. Since the final system will be for music listening, music signals were sampled. Short pieces of songs from three different artists (The Pixies, Sonic Youth, and Suzanne Vega) were sampled at 48kHz in stereo. The source of the music recordings were fairly obscure. None of the participants recognized the songs used. Thus they likely had no previous listening experience with the specific sounds used.

As in the previous tests, no discrimination of the listeners was made. The listeners would not have had any previous experience or practice with the test procedure. The subjects were not screened to insure good localizers.

The sounds were processed using HRTFs just like in test *A*. HRTFs at twelve locations were taken from Blauert's [Bla83] book: 0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, and 330°. The frequency response specifications are plotted in Figures 5.3–5.14 where $ang\{H(f)\} \equiv \angle H(f)$. Each sound was partitioned into either twelve or twenty-four sections. The length of the sections varied from sound to sound. The sections were either 0.5, 0.75, or 1.0 seconds long. Each section was processed with two separate HRTFs. The first HRTF was the position that the section started at, the second was the position that the section finished at. Each section moved through a 30° angle by mixing the two sections processed by the HRTFs. For example, the first section (assuming it would rotate clockwise) would be processed with the HRTFs $H_{h,90}(f)$ and $H_{h,120}(f)$. Then the sounds would be

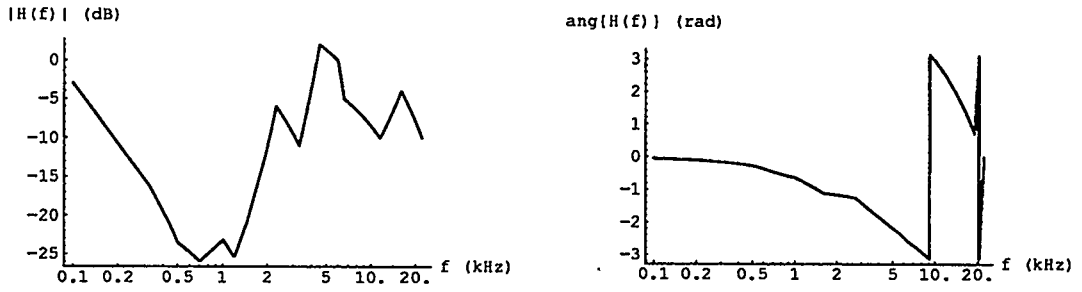

Figure 5.3: Magnitude response of $H_{e,0}(f)$.

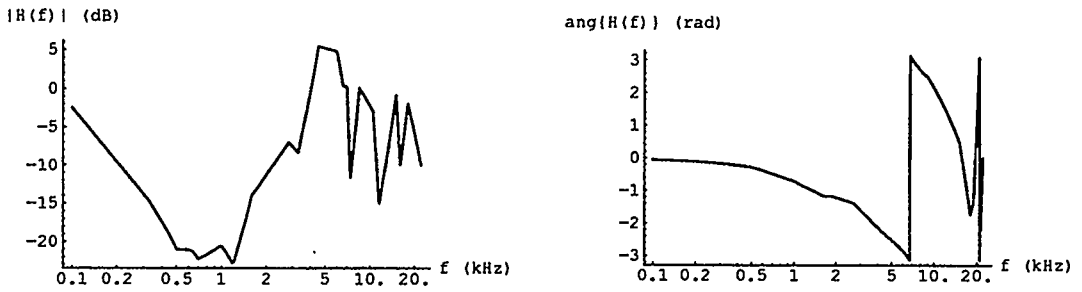

Figure 5.4: Magnitude response of $H_{e,30}(f)$.

mixed together. At first the $H_{h,90}(f)$ section would be mixed together at 100% and the $H_{h,120}(f)$ section at 0%. Then the mixture would linearly change throughout the length of the section until at the end the mixture was $H_{h,90}(f)$ at 0% and $H_{h,120}(f)$ at 100%. According to the summing localization effect discussed in section 1.8.1, this would make the sound appear to move between the first and the second position. Thus the first section of the sound moves through a 30° arc. This same procedure was then duplicated with the next section of the sound. Then all the sections were played one after each other to make the sound spin around the head. Figure 5.15 gives a pictorial representation of this. The rate at which the sound spins around the head is determined by the length of the sections. The shorter the section length, the faster the revolutions.

The total length of each sound was between 9 and 18 seconds. If the sound

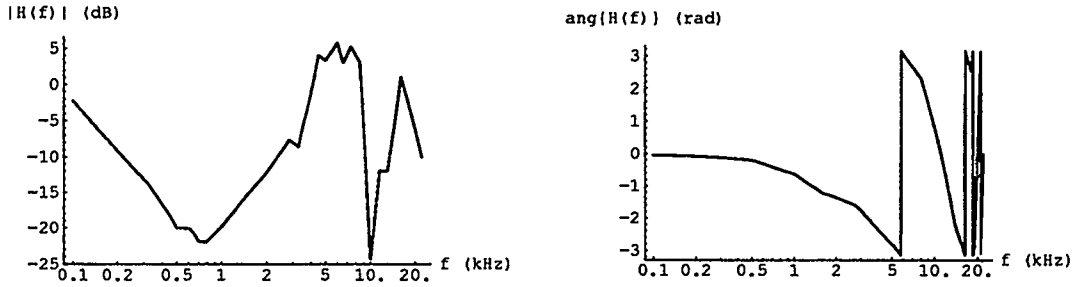

Figure 5.5: Magnitude response of $H_{e,60}(f)$.

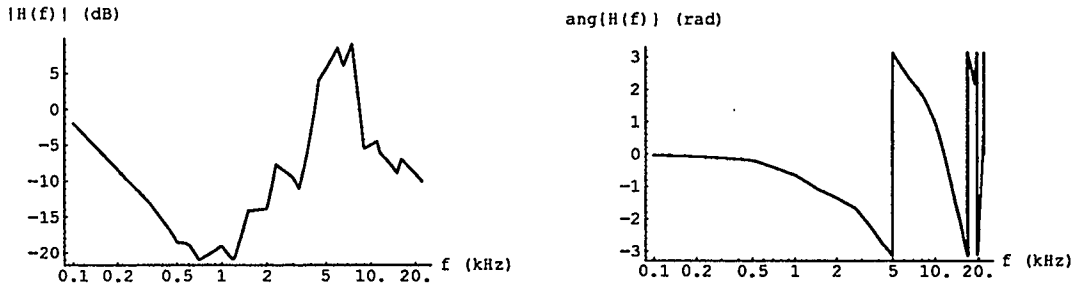

Figure 5.6: Magnitude response of $H_{e,90}(f)$.

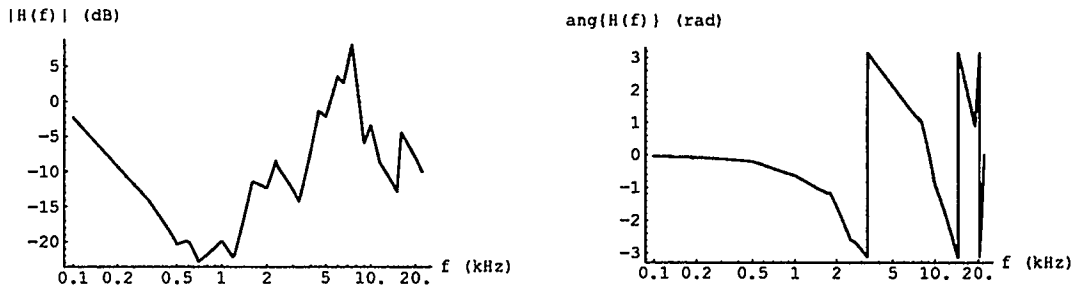

Figure 5.7: Magnitude response of $H_{e,120}(f)$.

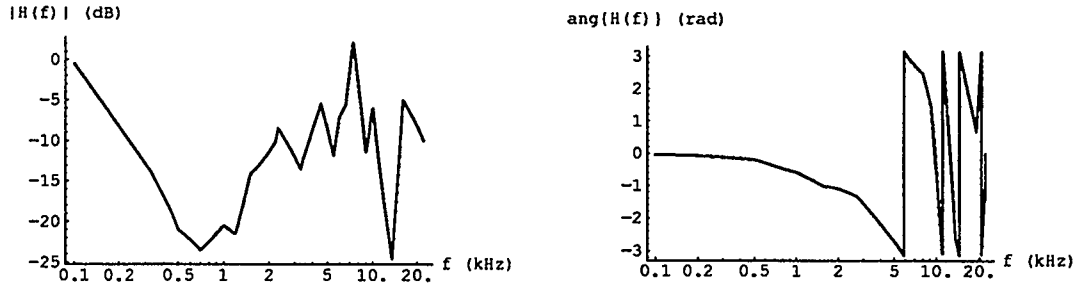

Figure 5.8: Magnitude response of $H_{e,150}(f)$.

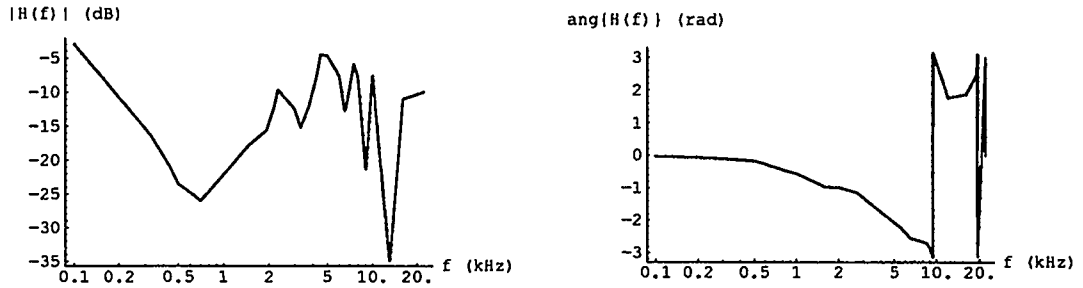

Figure 5.9: Magnitude response of $H_{e,180}(f)$.

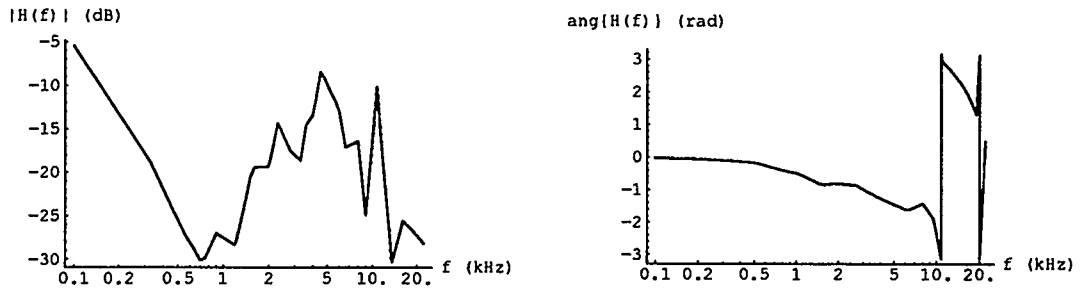

Figure 5.10: Magnitude response of $H_{e,210}(f)$.

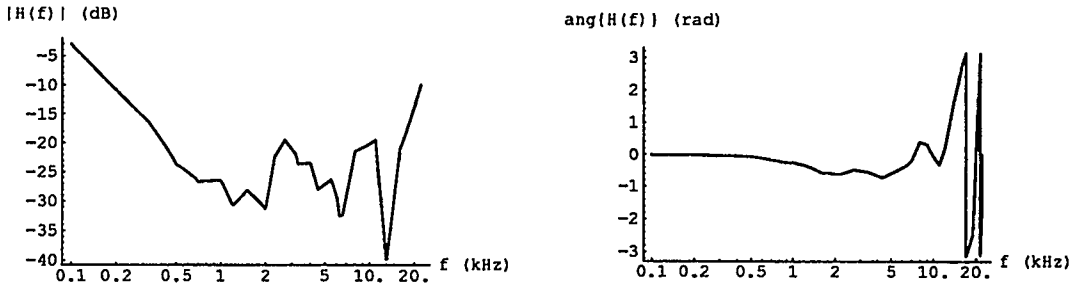

Figure 5.11: Magnitude response of $H_{e,240}(f)$.

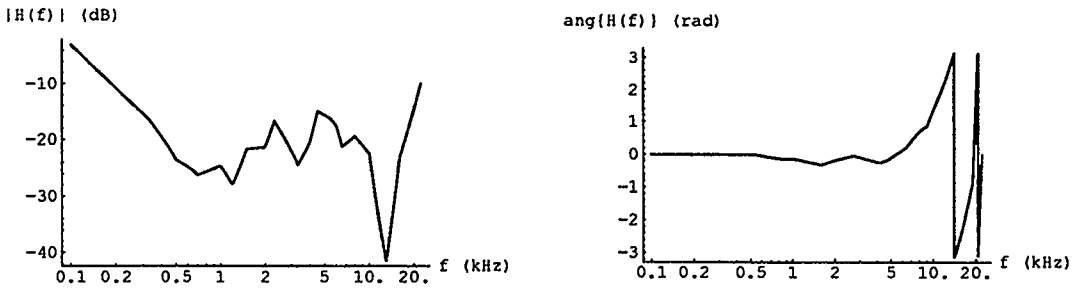

Figure 5.12: Magnitude response of $H_{e,270}(f)$.

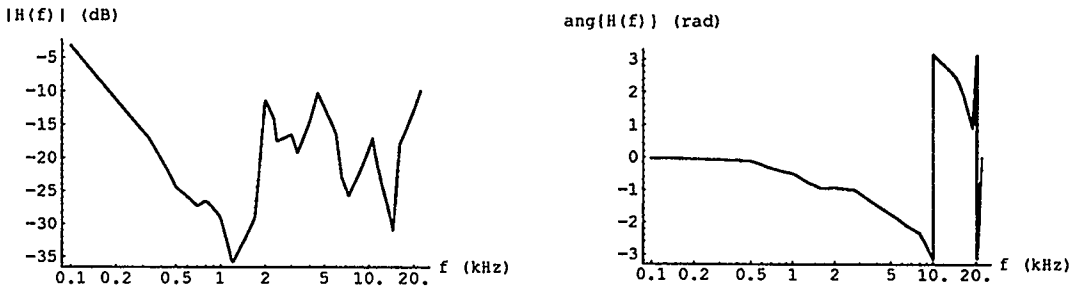

Figure 5.13: Magnitude response of $H_{e,300}(f)$.

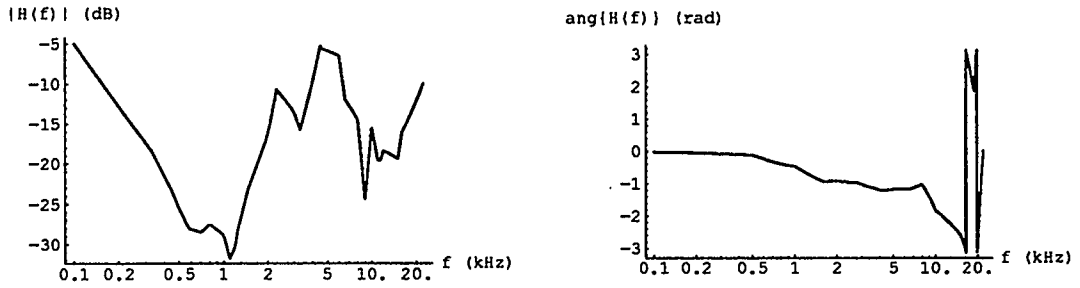

Figure 5.14: Magnitude response of $H_{e,330}(f)$.

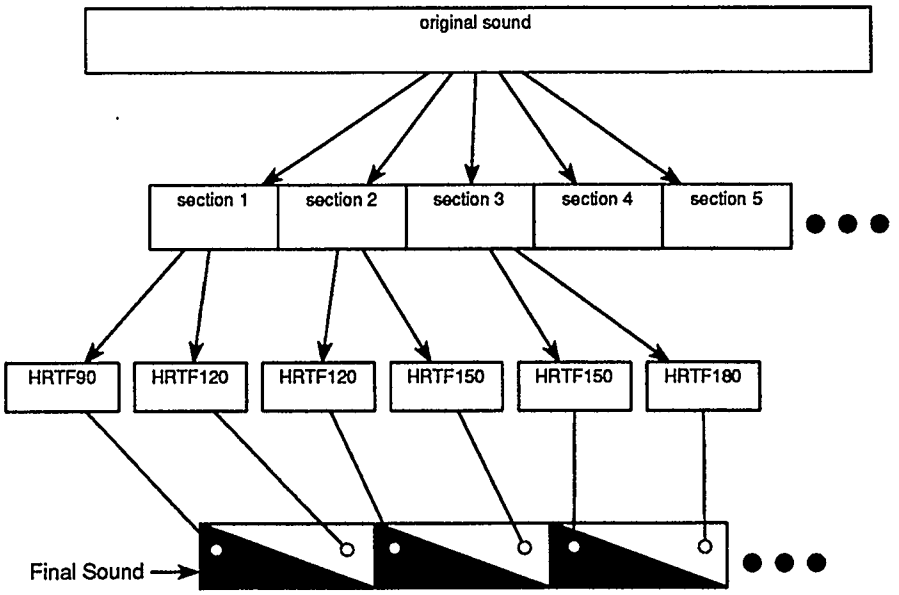

Figure 5.15: Pictorial representation of how a sound is rotated around the head in a clockwise direction.

section length was 0.5 seconds, the sound was divided into 24 sections and spun around the head twice. If the sections were 1.0 second long, the sound was divided into 12 sections and spun around the head once. If the sections were 0.75 seconds long, the sound was divided into either 12 or 24 sections and the sound was rotated once or twice around the head. The length of the sections was selected randomly. It should be noted that all the sounds started at 90° and ended at 90°.

A reverberated version of the sound was added randomly. The reverberation algorithm was described in Section 4.4. The reverberation might make the sound easier to locate because it sounded more natural. The experimental results should confirm or deny this.

To sum up, thirty-six sounds were created that varied in the direction (clockwise or counter clockwise), speed (0.5, 0.75, or 1.0 second sections), and reverberation (yes or no). The sounds were recorded onto a DAT tape and played for 9 subjects. Each sound was played twice without pause and then a slight pause indicated the next sound was to begin. The subjects were given written instructions telling them that thirty-six sounds would be played for them. Since the instructions were written down, all participants received the same instructions. They were informed that the sound would be moving around their heads and that they were required to determine whether it was going clockwise or counterclockwise. They then wrote their responses on a form provided.

5.3.1 Test Results

The percentage of correct decisions for each participant over the entire thirty-six sounds is given in Table 5.1.

The experiment shows good and bad locators. Subjects A, B, E and I were poor localizers. Subjects D, F, and G were good locators.

subject	*average*
A	55.6%
B	55.6
C	63.9
D	75.0
E	44.4
F	77.8
G	80.6
H	69.4
I	52.8
average	63.9

Table 5.1: Percentage correct for all sounds.

sounds	*average*
1–12	54.6%
13–24	67.6
25–36	69.4

Table 5.2: Average correct decisions for groups of twelve sounds.

Table 5.2 shows the average of correct decisions for the first, second and third set of twelve sounds. Since the percentage increases, it can be stated that learning took place. The effects of fatigue were observed at the end of the test. The last twelve sounds were divided into two groups: the last six and the second last six The second last group has a average correct guess of 75.9%, while the last group falls markedly to 63.0%. The subjects were seen to be bored and tired by the end of the test.

The best results occurred after learning and before fatigue. At this point, the average of correct responses was 75.9%. This is a very encouraging result after only ten minutes of adaptation.

The speed of the sounds effected the correct answer rate. Table 5.3 shows the effect of speed in the experiment. As the speed increases (section length gets shorter)

section length (seconds)	*average*
1.00	58.1%
0.75	65.0
0.50	70.0

Table 5.3: Average correct decisions for different section lengths.

the correct decision ratio goes up significantly. Thus, the rate of revolution has an effect on the experiment.

Reverberation does not have an effect on the subjects decision. The average of correct decisions with reverberation is 63.2% and without reverberation 64.7%. Wightman and Kistler's results predicted this outcome. The reverberation was added to achieve OHL because IHL occurred when the sounds were played in the unfamiliar anechoic environment. But since the subjects were given time to familiarize themselves with the environment, the reverberation was not required. In Wightman and Kistler's experiment the listeners experienced OHL in an anechoic environment because they were allowed to familiarize themselves with the environment. The reverberation may be retained in the localization system to aid OHL, but in theory is not necessary if the listeners can adapt to the environment.

The type of sound used might have also effected the decision. Some subjects reported that the music distracted them because they were more interested in hearing a lyric or listening to a melody than in localizing. The spectral properties of the sound would also effect its localization. Blauert [Bla70] shows how the frequency response of a sound can fool the senses into an improper localization by having certain localization frequency-bands boosted. The system used in test C is meant to be used with music. The variable effects of music selection were included in the experiment because they would be prevalent in the system's normal use.

Test C has showed that it is possible to teach people to localize in an artificial

environment using headphones. The percentage of correct localizations in the test was very encouraging. The system in part C allowed the listeners to localize sound better then they were able to in experiment A, especially after the subjects were given a chance to adapt to the environment. The most important point is that the subjects learned and that eventually they will be able to localize better. The only information given to the subjects about the sounds was a vague description of its movement. Once trained in this environment, a listener would be able to localize very efficiently. At some point the listener would be able to store the environment into their memory and not have to adapt to it any more.

The environment in test C consisted of twelve different channels defined by twelve different HRTFs; hence, the system used showed how a multi-channel system might be realized. A multi-channel system should contain a method of teaching the listener about the environment simulated and then HRTFs could be used to encode sounds into any number of locations.

Chapter 6

Conclusions

Headphones are an ideal method of controlling the sound heard by a listener, but headphones normally have a poor sound stage that is located inside the head. Recordings listened to on headphones produce a poor sound stage because these recordings are meant to be listened to in a free-field environment. The perceived sound stage of headphones was improved by simulating a free-field environment within the headphone environment. This was accomplished by using a method of controlling the location of auditory events. Several methods were investigated before a suitable one was discovered.

The first method used short echoes. This method assumed that sound passing through the pinna is echoed by the pinna's reflective surfaces. The echo time, which varies with the angle of sound incidence, tells the brain where the sound is located. It was found that an echo model does not simulate the effects of the pinna well enough to control auditory localization.

A better means of modeling the transfer function of the pinna is to measure its response directly. It was found that auditory localization could be controlled using these measurements, which are called Head Related Transfer Functions (HRTFs). However, the locations perceived had a large angular blur and the headphone's sound stage did not move outside of the listener's head.

When reverberation was added to a sound, its perceived location did move outside of the listener's head because reverberation makes the sound ear-adequate. When a sound is listened to in a normal (not a headphone) environment, the environment alters the sound. These alterations are a cue to the brain that sound is located outside the head. Reverberation duplicates some common environmental alterations, and

therefore makes the sound adequate for localization outside of the listener's head. A headphone system that simulated a free-field environment using HRTFs and reverberation was realized and tested. Listener's perceived that the system improved the sound stage, but the auditory localization blur was still large. A bit-serial hardware realization of the free-field simulator was constructed using programmable logic devices.

It was assumed that the method of controlling auditory localization used in the free-field system could be expanded to simulate multi-channel environments. As an experiment, HRTFs were used to try and locate auditory events in front of, and behind of, the listener. On average, the listeners perceived the locations as desired, but the variance of the results was large. A method achieving more consistent results was needed.

It was determined that in order for listeners to localize a sound, they must be familiar with the environment the sound is heard in. In normal (not headphone) listening, the subject adapts to the environment they are in by using their memory of previously encountered environments. Since most listeners have never encountered an artificial environment simulated on headphones, they would not be able to localize in an artificial environment unless they could adapt to it. People normally adapt to an environment using all their senses, but only the sense of hearing can be used to learn about a headphone environment. Is auditory information presented with headphones sufficient to allow adaptation?

Investigations showed that hearing is sufficient for headphone environment adaptation. A multi-channel headphone environment that could (theoretically) locate sounds around the listener in the horizontal plane was realized using HRTFs and reverberation. The sounds were rotated around the listener's head in the horizontal plane, and the listeners tried to perceive whether the sounds were rotated clockwise

or counter-clockwise. During the experiment, the listener's learned to localize the sounds more proficiently; therefore, they adapted to the environment.

In essence, a simulated headphone environment is a virtual world. In this world, the subject could not see, smell, move, touch or taste; they could only hear. Hearing in this virtual world was very similar to hearing in the real world. Simulated pinnas encoded information about the location of an event into the frequency spectrum of a sound, and the environment altered the sound in familiar ways (reverberation). In order for the listener to perceive this world, they had to learn to recognize these cues. They were generally the same as "real-world" audio cues, but subtile differences existed that had to be learned. For example, the simulated pinnas were not identical to the listeners real pinna. Before these subtile differences were learned, the localizations were blurred (HRTF presentation without adaptation had a large angular blur).

Further research into the ability of the listener to adapt to the simulated headphone environment is necessary. It has not been determined whether a listener can memorize a headphone environment, and thus require no further adaptation. If a listener can not, then training sounds would have to be a regular part of any listening programme.

Bibliography

[Asc63] V. Aschoff. Über das räumliche hören [on spatial hearing]. *Arbeitsgem. f. Forschung Nordrhein-Westfalen*, 138:7–38, 1963. Westdeutscher Verlag, Köln. cited in [Bla83].

[Bat67] D. W. Batteau. The role of the pinna in human localization. *Proceedings of the Royal Society of London (series B)*, 168:158–180, 1967.

[Bau61] B. B. Bauer. Stereophonic earphones and binaural loudspeakers. *Journal of the Audio Engineering Society*, 9(2):148–151, 1961. Westdeutscher Verlag, Köln. cited in [Bla83].

[BB77] Robert A. Butler and Krystyna Belendiuk. Spectral cues utilized in the localization of sound in the median sagittal plane. *Journal of the Acoustical Society of America*, 61(5):1264–1269, 1977.

[BL78] J. Blauert and P. Laws. Verfahren zur orts- und klanggetreuen simulation von lautsprecherbeschallungen mit hilfe von kopfhörern [true simulation of loudspeaker sound reproduction while using headphones]. *Acustica*, 29:273–277, 1978.

[Bla83] Jens Blauert. *Spatial Hearing.* MIT Press, 1983.

[Bla70] Jens Blauert. Sound localization in the median plane. *Acustica*, 22:205–213, 1969/70.

[Car66] Harvey A. Carr. *An Introduction to Space Perception.* Hafner Publishing Company (New York), 1966. Facsimile of 1935 edition.

[Cry92] Crystal Semiconductor Corporation. *Volume I Data Book, Analog/Digital Conversion IC's*, 1992.

[EA90] Mark A Ericson and Jeffrey R Agnew. A comparison of localization performance with two auditory cue synthesizers. *Proceedings of the IEEE 1990 Aerospace Electronics Conference*, 2(NAECON):749–754, 1990.

[EMWF88] Frederic L. Wightman Elizabeth M. Wenzel and Scott H. Foster. Development of a three-dimensional auditory display system. *SIGCHI Bulletin*, 20(2):52–57, 1988.

[FF67] H. Geoffrey Fisher and Sanford J. Freedman. Localization of sound during simulated unilateral conductive hearing loss., October 1967. cited in [FF68].

[FF68] H. Geoffrey Fisher and Sanford J. Freedman. The role of the pinna in auditory localization. *Journal of the Acoustical Society of America*, 8:15–26, 1968.

[Gra89] Peter Graumann. *DFIRST User's Guide.* Department of Electrical and Computer Engineering, University of Calgary, 1989.

[GT92] Peter Graumann and Laurence Turner. *TRANS User's Guide.* Department of Electrical and Computer Engineering, University of Calgary, 1992.

[Hau69] B. G. Haustein. Hypothesen über die einohrige entfernungswahrnehmung des menschlichen gehörs [hypotheses about the perception of distance in human hearing with one ear]. *Hochfrequenztech. u. Elektroakustik*, 78:46–57, 1969. cited in [Bla83].

[KGST92] Ray Kacelenga, Peter Graumann, Mike Svihura, and Laurence Turner. Nomad, 1992. Department of Electrical and Computer Engineering, University of Calgary.

[KTBW89] B. Kish, L. E. Turner, J. M. Bauer, and R. Wheatley. *LOGSIM Users Guide.* Department of Electrical and Computer Engineering, University of Calgary, 1989.

[LHC90] Jack M. Loomis, Chick Hebert, and Joseph G. Cicinelli. Active localization of virtual sounds. *Journal of the Acoustical Society of America*, 88(4):1757–1764, 1990.

[LM88] Richard F. Lyon and Carver Mead. An analog electronic cochlea. *IEEE Transaction on Acoustics, Speech, and Signal Processing*, 36(7):1119–1133, 1988.

[MD89] Gary S. Kendall William L. Martens and Shawn L. Decker. Spatial reverbation: Discussion and demonstration. In Max V. Mathews and John R. Pierce, editors, *Current Directions in Computer Music Research*, chapter 7. MIT Press, 1989.

[Moo79] James A. Moorer. About this reverberation business. *Computer Music Journal*, 3(2):13–28, 1979.

[Moo82] Brian C. J. Moore. *An Introduction to the Psychology of Hearing.* Academic Press Inc. (LONDON) Ltd., 1982.

[OS89] Alan V. Oppenheim and Ronald W. Schafer. *Discrete-Time Signal Processing.* Prentice-Hall, 1989.

[Pei63] Peiper. Unknown. *Unknown*, ?:?, 1963. Reference cited on pg. 178 in [Moo82] but was omitted from the articles references.

[Ple74] G. Plenge. On the differences between localization and lateralization. *Journal of the Acoustical Society of America*, 56(3):944–951, 1974.

[Ray07] Lord Rayleigh. On our perception of sound direction. *Phil. Mag.*, sixth series(13):214–232, 1907. cited [EMWF88].

[Sch62] M. R. Schroeder. Natural sounding artificial reverbation. *Journal of the Audio Engineering Society*, 10(3):219–223, 1962.

[SGK76] N. Sakamoto, T. Gotoh, and Y. Kimura. On "out-of-head localization" in headphone listening. *Journal of the Audio Engineering Society*, 24(9):710–715, 1976.

[ST68] E. A. G. Shaw and R. Teranishi. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *Journal of the Acoustical Society of America*, 44(1):240–249, 1968.

[WHW74] Donald Wright, John H. Hebrank, and Blake Wilson. Pinna reflections as cues for localization. *Journal of the Acoustical Society of America*, 56(3):957–962, 1974.

[WK89a] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. I: Stimulus synthesis. *Journal of the Acoustical Society of America*, 85(2):858–867, 1989.

[WK89b] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. II: Psychophysical validation. *Journal of the Acoustical Society of America*, 85(2):868–878, 1989.

[Xil91] Xilinx Inc. *The XC4000 Data Book*, 1991.

[ZC81] P. M. Zurek and W. W. Clark. Narrow-band acoustic signals emitted by chinchilla ears after noise exposure. *Journal of the Acoustical Society of America*, 70:446, 1981. cited in Nielson L., M. A. Mahowald and C. Mead, SEEHEAR. In C. Mead, editor, *Analog VLSI and Neural Systems,* pages 207–227. Addison-Wesley, 1989.

[Zur87] P. M. Zurek. The precedence effect. In William A. Yost and George Gourevitch, editors, *Directional Hearing*, chapter 4. Springer-Verlag, 1987.