

University of Windsor

Scholarship at UWindor

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

2001

Auditory model implementation for speech application.

Jie. Zhang

University of Windsor

Follow this and additional works at: <https://scholar.uwindsor.ca/etd>

Recommended Citation

Zhang, Jie., "Auditory model implementation for speech application." (2001). *Electronic Theses and Dissertations*. 1246.

<https://scholar.uwindsor.ca/etd/1246>

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

AUDITORY MODEL IMPLEMENTATION FOR SPEECH APPLICATION

by

Jie Zhang

A Thesis

Submitted to the Faculty of Graduate Studies & Research
through the Faculty of Engineering – Electrical & Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada

December, 2000



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-62308-4

Canada

© 2000 Jie Zhang

All Rights Reserved. No part of this document may be reproduced, stored or otherwise retained in a retrieval system or transmitted in any form, on any medium or by any means without the prior written permission of the author.

ABSTRACT

The motivation for investigating spectral analysis methods that are physiologically based is to gain an understanding of how the human auditory system processes speech, so as to be able to design and implement robust, efficient methods of analyzing and representing speech. In this thesis the cochlear filter bank is modeled by a lowpass filter cascaded model. In order to model the effect of a notch in the frequency response approximately one octave below the center frequency a second filter is followed after each notch/ resonant filter in the cascaded model. We also employed the genetic algorithms to optimize the filter parameters. The filters in the model are implemented using four-multiply normalized ladder filter form. In this thesis we also explored the ensemble interval histogram (EIH) representation in the speech recognition. This representation models the properties of the post auditory-nerve nuclei in the human auditory system. The simulation results show that the overall recognition rate is increased a little.

To China, the land I could never forget.

To Canada, the land I call home.

ACKNOWLEDGMENTS

I wish to thank Dr. H. K. Kwan, my thesis supervisor, for introducing me the subject of this thesis, for his intellectual advice and guidance on the thesis and other academic matters during the period of my master degree program. I would like to acknowledge Dr. Kwan for suggesting me to adopt the combination of the Lyon's cascaded filter cochlear model, the Ensemble Interval Histogram (EIH) representation, LVQ1 and LVQ3 in the formation of the auditory and speech recognition system in this thesis. In particular, I wish to acknowledge Dr. Kwan's suggestions to me on the use of the normalized ladder filter for the realization of the cascaded 2nd-order filters for the cochlear model, and on the use of genetic algorithms for the cascaded 2nd-order filter design based on a scaled version of the cat's nerve-fiber responses. I also would like to acknowledge Dr. Kwan for providing his created set of 50 in-vehicle commands for uses in the experiments of this thesis. Dr. Kwan's comments and suggestions were particularly useful and benefited every stage of this thesis.

I also would like to thank my department reader, Dr. Xiang Chen, and my external reader, Dr. Walid Saba, for serving as my committee members and for their helpful comments with regard to the thesis.

On a personal note, I would like to acknowledge and thank the assistance of all my colleagues in the ISPLab: Wayne Chiang, Tracy Li, Hui Ping, Walter Jin, Liang Tao, Halima El-Khatib, Ming Wang and Xiao Feng Dong. Their strong work ethic and sincere concern for others will be a lasting image from my time at the University of Windsor.

My family also deserves to be credited for the strong supports they have given me throughout my university career. Finally, a great deal of gratitude and thanks goes to my wife, Janet. Her patience, understanding, and support over the last two years have been monumental and will not be forgotten.

Contents

Abstract.....	v
Acknowledgments.....	vii
Chapter 1 The Fundamentals of Cochlear Filter Design.....	1
Section 1 Introduction.....	1
1.1 Motivation of the Thesis.....	1
1.2 Organization of the Thesis.....	2
Section 2 Physiological Fundamentals.....	3
2.1 The Physical Structure of Human Ear.....	3
2.2 The Cochlear Structure.....	4
Section 3 Cochlear Filter Bank.....	8
3.1 Low-pass Filter Cascaded Model.....	8
3.2 Transmission-line Model.....	10
3.3 Phenomenological Filter Model.....	19
3.4 The Problems in the Existing Cochlear Model and Objective of Our Research.....	22
Section 4 Genetic Search Approach.....	23
4.1 The Outline of the GA.....	23
4.2 Reproduction.....	25
4.3 Elitism.....	25
4.4 Crossover.....	26

4.5	Mutation.....	27
4.6	GA for IIR Filter Optimization.....	28
Section 5	Lattice Form Realization.....	30
5.1	Problems in the Traditional Canonic Direct-form.....	30
5.2	Principles of the Four-multiply Normalized Ladder Filter.....	31
5.3	Regalia and Mitra's Method.....	31
5.4	First-order Lattice Filter.....	32
5.5	Second-order Lattice Filter.....	33
5.6	The Four-multiply Markel-Gary Normalized Ladder.....	34
Section 6	Ensemble Interval Histogram (EIH) Representation.....	36
6.1	Level-crossing.....	37
6.2	Measurement of Synchrony and Instantaneous Rate Across the Simulated Fibers.....	37
Chapter 2	Cochlear Filter Model Description.....	38
Section 1	Introduction.....	39
Section 2	Outer/middle Ear Combined Bandpass Filter.....	40
Section 3	Filterbank.....	42
3.1	Definitions.....	42
3.2	Bandwidth.....	43
3.3	Center Frequency.....	44
3.4	Filter Q Values.....	46
3.5	Filter Gain.....	47
Section 4	Second Filter.....	49
Section 5	Normalized Lattice Form Realization of 2nd Order Digital Filter....	50

Section 6	Filter Optimization Using GA.....	52
Section 7	Half Wave Rectification.....	53
Section 8	Nonlinear Stages.....	54
Section 9	Summary.....	59
Chapter 3	Ensemble Interval Histogram Model.....	60
Section 1	Introduction.....	61
Section 2	Cochlear Filterbank.....	62
Section 3	Level Crossing Detection.....	62
Section 4	Interval Histogram.....	62
Section 5	Summation of the EIH from All Channels	64
Section 6	The Choice of Bin Allocation over Frequency Range.....	65
Section 7	The Choice of the Window Length.....	67
Chapter 4	Experiment Results.....	68
Section 1	Introduction.....	68
Section 2	Filterbank Simulation Results and Analysis.....	69
Section 3	EIH Simulation Results and Analysis.....	77
Section 4	Speech Recognition Experiment Results.....	87
4.1	Introduction.....	87
4.2	Signal Conditions.....	87
4.3	Analysis Methods.....	87
4.4	Experiment Results and Analysis.....	88

Chapter 5 Conclusions and Future Works.....	95
Section 1 Introduction.....	95
Section 2 Cochlear Model.....	95
2.1 Filterbank.....	96
2.2 Second Filter.....	96
2.3 Half Wave Rectifier and Nonlinear Stage.....	96
2.4 Four-multiply Normalized Ladder Filter Implementation.....	97
2.5 Filter Optimization Using Genetic Algorithms.....	97
Section 3 Ensemble Interval Histogram.....	97
Section 4 Evaluation.....	98
Section 5 Future works.....	98
REFERNCES.....	99
VITA AUCTORIS.....	100

List of Tables

Table 1.1	24
Table 1.2	25
Table 1.3	28
Table 2.1	52
Table 2.2	56
Table 4.1	93

List of Figures

Figure 1.1	3
Figure 1.2	4
Figure 1.3	5
Figure 1.4	7
Figure 1.5	8
Figure 1.6	10
Figure 1.7	12
Figure 1.8	13
Figure 1.9	14
Figure 1.10	15
Figure 1.11	18
Figure 1.12	20
Figure 1.13	21
Figure 1.14	31
Figure 1.15	32
Figure 1.16	33
Figure 1.17	34
Figure 1.18	35
Figure 1.19	36
Figure 2.1	39
Figure 2.2	41
Figure 2.3	44

Figure 2.4	45
Figure 2.5	46
Figure 2. 6	48
Figure 2.7	50
Figure 2.8	53
Figure 2.9	54
Figure 2.10	56
Figure 2.11	57
Figure 3.1	61
Figure 3.2	63
Figure 3.3	64
Figure 3.4	66
Figure 3.5	67
Figure 4.1	70
Figure 4.2	78
Figure 4.3	83
Figure 4.4	87
Figure 4.5	89

List of Abbreviations

GA	genetic algorithm
EIH	ensemble internal histogram
OHC	outer hair cell
IHC	inner hair cell
BM	basilar member
RL	reticular lamina
HB	hair bundle
TM	tectorial membrane
DC	Deiter cell
PC	pillar cells
CF	characteristic frequency
AGC	automatic gain control
HWR	half-wave rectifier
SV	scala vestibuli
ST	scala tympani
CP	cochlear partition
ERB	equivalent rectangular bandwidth
RCS	radial cross section
MBPNL	multiple band-pass non-linearity
IIR	infinite impulse response
SNR	signal to noise ratio

CHAPTER 1

The Fundamentals of Cochlear Filter Design

Section 1 Introduction

Section2 Physiological Fundamentals

Section 3 Cochlear Filter Bank

Section 4 Genetic Search Approach

Section 5 Lattice Form Realization

Section 6 Ensemble Interval Histogram (EIH) Representation

1. Introduction

1.1 Motivation of the Thesis

The cochlea is a part of the inner ear in which acoustic signals are converted to the neural code that conveys auditory information into the brain. Modeling the function of the cochlear has been an active area of research with the development of the digital signal processing. The premise of cochlear modeling researches is to well understand human auditory perception processes.

Depending on these understandings to the cochlea, more robust, efficient methods of analyzing and representing speech methods can be found because it is generally realized that the better we understand the signal processing in the human auditory system, the closer we will be able to design a system that can truly understand meaning of speech.

In the cochlear model a model of the cochlea and the hair cell transduction consists of a filter bank that models the frequency selectivity at various points along a simulated basilar membrane, and a nonlinear processor for converting the filter bank output to neural firing patterns along a simulated auditory nerve. Such a model is described in this thesis and is called the ensemble internal histogram (EIH) model [Ghitza 94].

1.2 Organization of the Thesis

In following sections of this chapter we will introduce some backgrounds of auditory model, which include physiological fundamentals of the cochlear model, survey of cochlear filter bank models, genetic algorithms (GA) for filter optimization, lattice form filter realization, and EIH model.

In chapter 2 we will discuss the implementation of the lowpass cascaded model. We will also discuss filter bank coefficient optimization using genetic algorithms and filter realization using lattice form.

In chapter 3 we will describe the ensemble internal histogram (EIH) representation in the cochlear model.

In chapter 4 we will present some simulation results using EIH model and analysis.

In last chapter we will gives the conclusions and some directions of the future works

2. Physiological Fundamentals

2.1 The Physical Structure of Human Ear

The physical structure of the human ear is shown in figure 1.1. It consists of three distinct parts: the *outer ear*, the *middle ear*, and the *inner ear*. The outer ear comprises the *pinna* and the *external canal*. Sound waves are guided through the outer ear to the middle ear, which comprises the eardrum and a mechanical transducer that includes the three small bones: *hammer*, the *anvil* and the *stirrup*. When air pressure in front of the eardrum increases, the eardrum is pushed inward, moving the hammer, anvil, and stirrup. The footplate of the stirrup covers the oval window of the cochlea and the movement of the stapes initiates a pressure wave in the cochlear fluid [Allen and Neely 92].

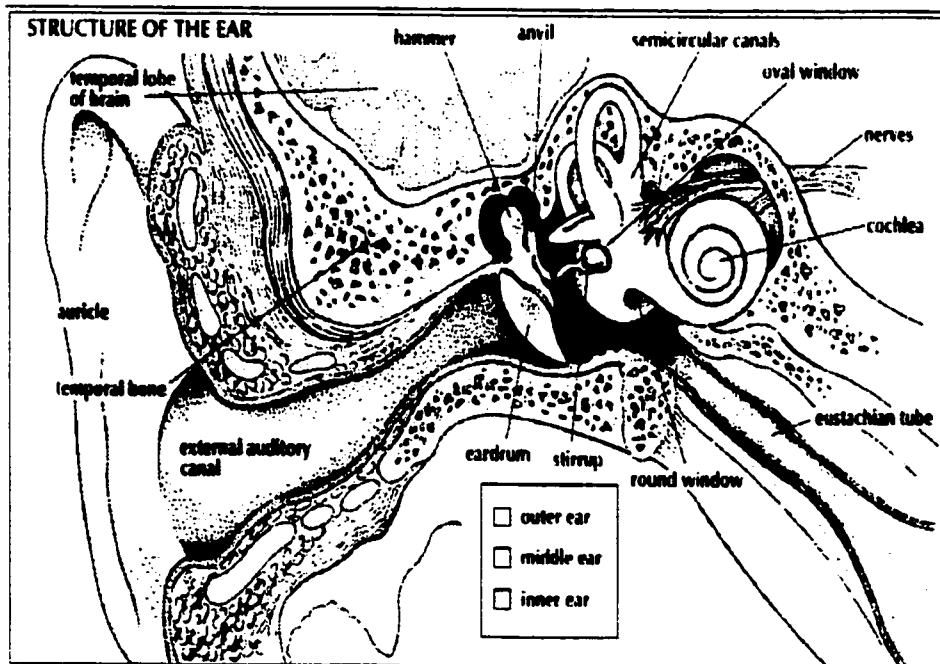


Fig 1.1 Physical structure of the human ear [Ghitza 94]

2.2 The Cochlear Structure

The cochlea is the portion of the inner ear devoted to hearing. It is a spiraling, fluid-filled tunnel embedded in the temporal bone. The mid-modiolus section (figure 1.2) shows the coiling of the cochlear duct, the scala vestibuli and the scala tympani. The red arrow is from the oval window, the blue arrow points to the round window. Within the cochlea, fluid-borne mechanical signals are transformed into the neural code delivered by the auditory nerve to the brain. Acoustic signals that enter the fluid-filled cochlear chambers propagate in a dispersive manner along the cochlear partition. The partition, which spans the length and width of the cochlea, consists of the basilar membrane, tectorial membrane, and organ of Corti [Allen and Neely 92].

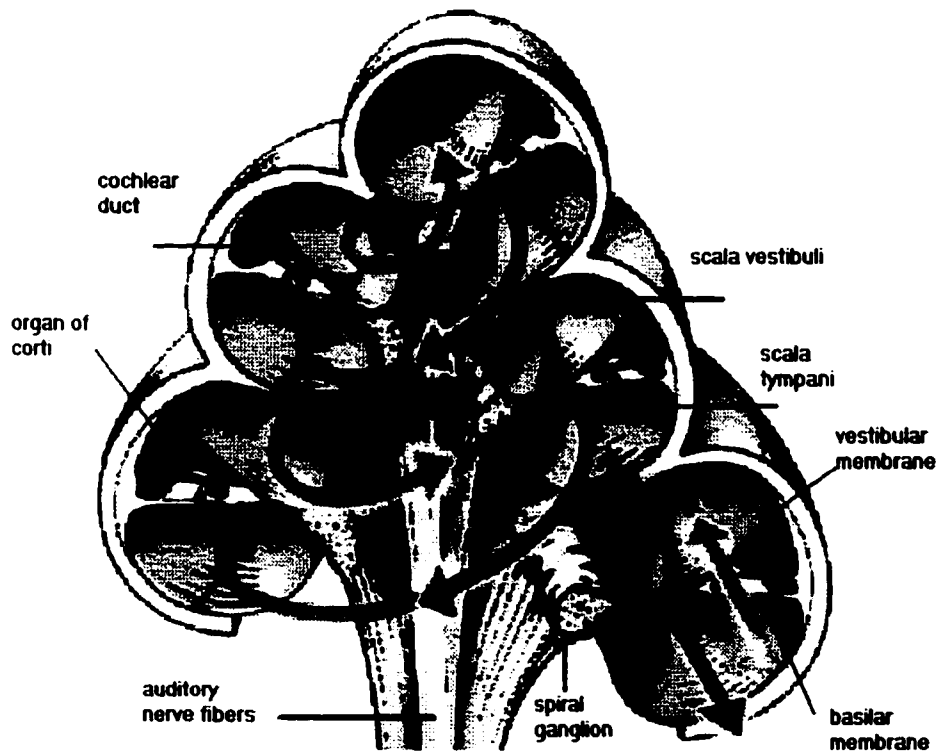


Fig 1.2 Mid-modiolus Section of the Cochlea [Allen and Neely 92]

Figure 1.3 illustrates the anatomical structure of the organ of Corti. The organ of Corti is a collection of cells, including the sensory hair cells, which sit on the basilar membrane. The outer hair cells (OHCs) and inner hair cells (IHCs) are positioned between the basilar member (BM) and reticular lamina (RL). The tips of the hair bundles (HBs) of OHC are embedded in the tectorial membrane (TM). The base of the OHC rests on the cup of the Deiter cell (DC). The IHCs and OHCs are separated by rigid pillar cells (PCs). The HB of IHC does not touch the TM and are displaced by viscous fluid drag [Allen and Neely 92].

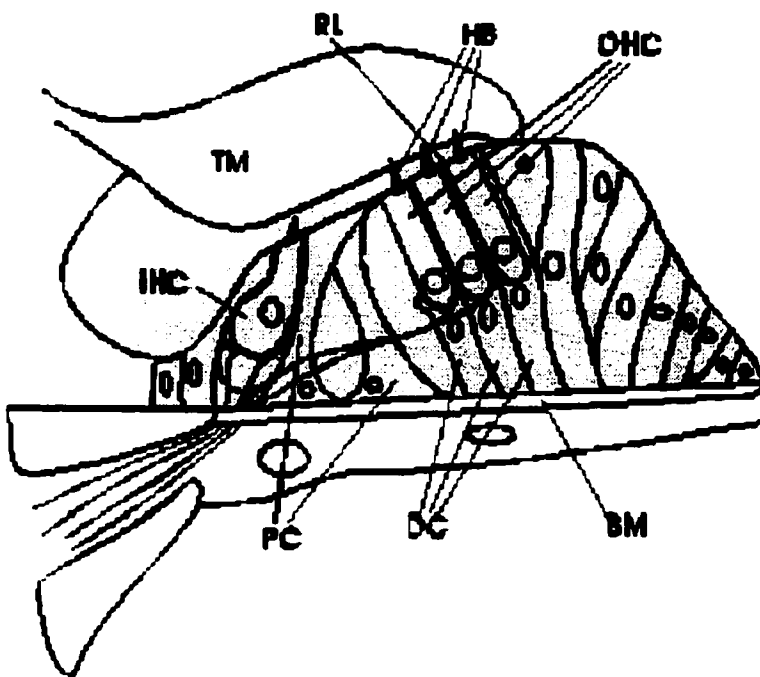


Fig. 1.3 Diagram of the Anatomical Structure of the Organ of Corti [Neely 93]

The IHCs are connected to the auditory nerve by afferent fibers, which deliver neural signals to the brain. IHCs are mechano-electric transducers that convert displacement of HB, which is due to shearing movement between the TM and RL (due to the basilar movement), into a receptor

current change in the cell. This is done by mechanical gating of ion channel that must be located in the HB.

The OHCs, on the other hand, are innervated by efferent nerve fibers, which receive neural signals from the brain. OHCs act as electro-mechanic transducers by converting voltages across their cell membranes into their length changes. This capability is important to cochlear non-linearity.

The human cochlea is believed to contain approximately 4000 IHCs and 12,000 OHCs, with four cells radially abreast and spaced every 10 microns along the BM's length. The TM lies on top of the organ of Corti and is attached, at its inner edge, to the bony spiral limbus. A thin fluid space (4 to 6 microns) lies between these two surfaces, which shear as the BM moves up and down.

The mechanical displacement of the BM, at any given place, can be viewed as the output of a band-pass filter with a given resonant frequency called characteristic frequency (CF). This filtering characteristic separates the various frequency components of the signal with a good signal-to-noise ratio. Moreover, each filter has its own dynamic range compression. This non-linearity makes the frequency response of each filter dependent on the level of the acoustic signal.

All mammalian cochleae appear to function according to the same basic principles; however, the effective frequency range differs among species. For example, the range of audible frequencies is about 20 Hz to 16 kHz in the human cochlea and about 100 Hz to 40kHz in the cat cochlea. The human basilar membrane is about 35 mm long, while the cat basilar membrane is about 25 mm.

A neural tuning curve is a “threshold of hearing” curve for a single neuron, using a sine-tone stimulus. Figure 1.4 shows a set of neural tuning curves. The tuning seen in neurons of the auditory nerve resembles filtering by an electro-mechanical filter. So the objective of designing cochlear models is to reproduce neural tuning curve. If the filter is linear then it is possible to find the magnitude of the transfer function by inverting the neural tuning curve.

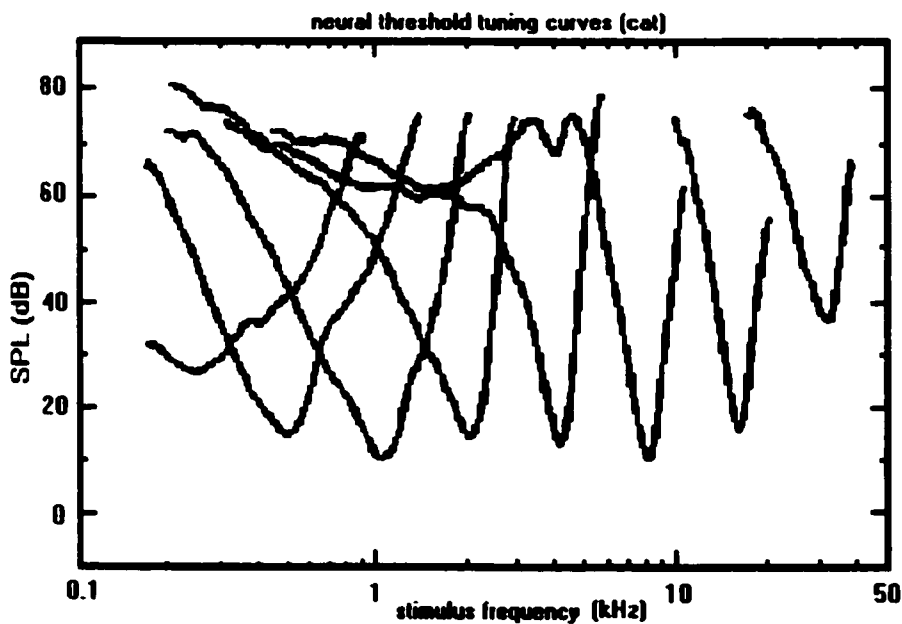


Fig. 1.4 Set of Tuning Curves [Allen and Neely 92]

3 Cochlear Filter Bank

3.1 Low-pass Filter Cascaded Model

In this structure the cochlear model is split into a large number of discrete sections. Each section can be model with a simple linear transfer function. The output at a specific location of the cochlear partition is the cascaded output of all the filter sections that precede it. These models include Lyon's model [Slaney and Lyon 90], Kates' model [Kates 93], Zhao's cochlear filter [Zhao 96], and AMRL's cochlear filter.

The structure of Lyon's model is shown in Figure 1.5. The cochlea model described by Lyon [Lyon 90] combines a series of notch filters that model the traveling pressure waves with resonators to model the conversion of pressure waves into basilar membrane motion or velocity.

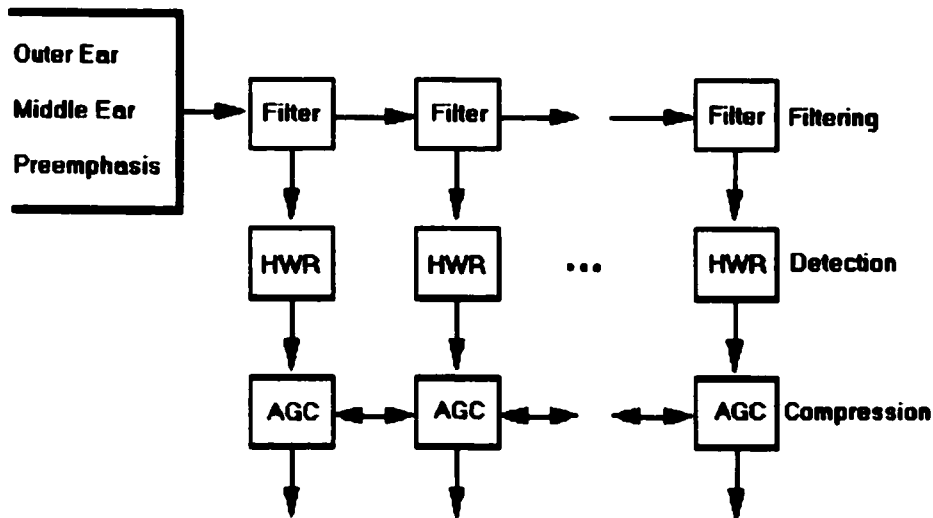


Fig. 1.5 The Structure of low-pass filter cascaded model [Slaney 88]

At each point in the cochlea the acoustic wave is filtered by a notch filter. Each notch filter operates at successively lower frequencies so the net effect is to gradually low pass filter the acoustic energy. An additional resonator (or bandpass filter) picks out a small range of the traveling energy and models the conversion into basilar membrane motion. It is this motion of the basilar membrane that is detected by the inner hair cells. A combination of a notch and a resonator is called a stage.

The output of each filter stage is a bandpass representation of the original audio signal. Each of these bandpass signals is passed through a half-wave rectifier and then through several AGC. The half-wave rectifier models the detection nonlinearity of the hair cells, providing a non-negative output that can be used to represent neural response. The half-wave rectification provides a crude energy measure in the signal. Each AGC stage is implemented as a variable gain that tries to keep the output of the AGC stage from exceeding a fixed level. In general the gain will be between zero and one. To model the masking effects of the ear, each stage of the AGC combines the bandpass outputs from the current channel plus its nearest neighbors. Since all channels are coupled one channel can affect all channels in the filter bank although the effect will decay exponentially with distance.

From model example frequency response (figure 1.6) and its structure analysis we can find that this kind of model accurately describes the total mechanical effects of the cochlea, such as wave propagation, frequency selection, nonlinearity of cochlea, masking and adaptation of the ear. It needs low computation burden comparing to other model. However, due to its simplicity it makes no attempt to accurately model each individual component.

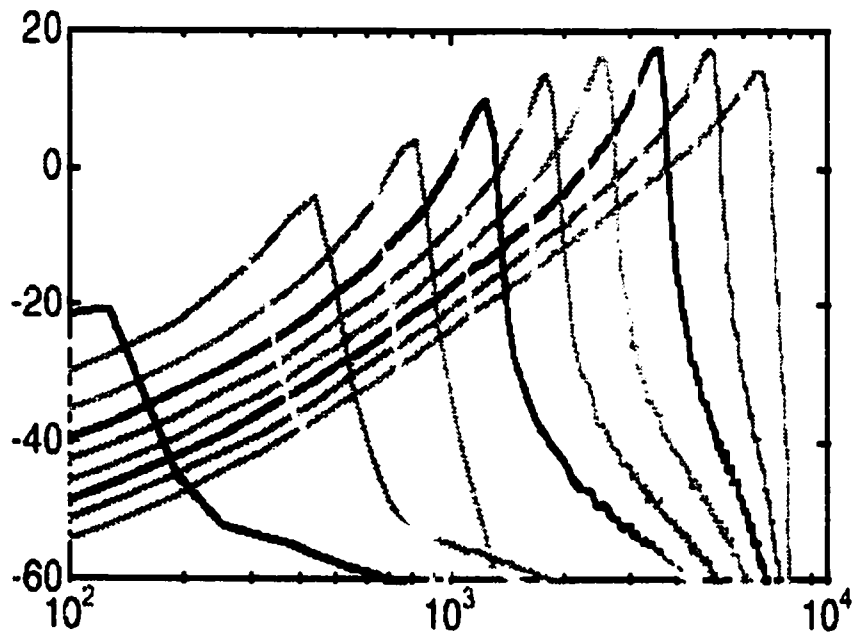


Fig. 1.6 The Example Frequency Response of the Low-pass Filter Cascaded Model [Slaney 88]

3.2 Transmission-line Model

In transmission-line models the wave propagation in the cochlea is simulated as the response of a transmission line with distributed parallel and series impedance that varies continuously with distance along the basilar membrane. Examples of these models are Allen's model [Allen and Neely 92], Neely's model [Neely 93], Deng's model, Giguere and Woodland's model. These models are described based on a one-dimensional low-frequency approximation to the cochlear wave equation.

The approach of Neely's model is described by several different models, i.e. macromechanics, middle ear, micromechanics, outer hair cell, inner hair cell, which will be described respectively below [Neely 93].

A. Macro-mechanics

The macro-mechanics refers to the mechanics of the fluid in scala vestibuli (SV) (upper compartment) and scala tympani (ST) (lower compartment) and interaction of the fluid with its boundaries at the stapes, round window, helicotrema, and cochlear partition (CP) (middle compartment). To simplify the macromechanics we assume that the cochlea is divided into a large number of thin slices by cutting it perpendicular to the x axis. Since the cochlear is spiraled, these slices are referred to as radial cross section (RCS). Each RCS has a thickness δ , which may vary as its location. The RCSs are mechanically coupled to adjacent sections only through the fluid compartments (SV and ST); any longitudinal coupling directly through the CP is ignored.

Figure 1.7 summarizes the representation of the macromechanics in these models for a single RCS at a distance x from the stapes. The combined mass of the SV and ST is viewed as a single mass element $M_f(x)$, i.e.:

$$M_f(x) = 2\rho\delta/A_c(x) \quad (1.1)$$

where ρ is the fluid density and $A_c(x)$ is the 'average' cross-sectional area of SV and ST.

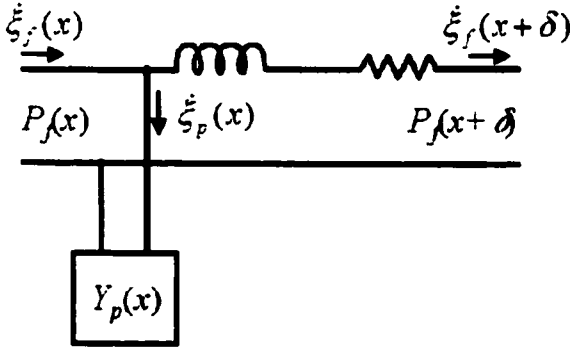


Fig. 1.7 Summary of the Cochlear Macromechanics for One RCS [Neely 93]

The damping parameter $R_f(x)$ is chosen as required for proper cochlear input impedance at low frequencies. The total acoustic admittance for the fluid in a single RCS is:

$$Y_f(x) = [R_f(x) + i\omega M_f(x)]^{-1} \quad (1.2)$$

$$\xi_f(x+\delta) = Y_f(x)[P_f(x) - P_f(x+\delta)] \quad (1.3)$$

$$\xi_p(x) = Y_p(x)P_f(x) \quad (1.4)$$

$$\xi_p(x) = \xi_f(x) - \xi_f(x+\delta) \quad (1.5)$$

$$[Y_p(x) + Y_f(x-\delta) + Y_f(x)]P_f(x) = Y_f(x-\delta)P_f(x-\delta) + Y_f(x)P_f(x+\delta) \quad (1.6)$$

where ξ_f , ξ_p are displacement of this section of RCS and CP respectively, and P_f , P_p are fluid pressure of this section of RCS and CP respectively.

The RCS at the apical boundary $x = L$ is terminated by a damping element R_h , which represents fluid passage through the helicotrema:

$$[R_h + Y_f(x - \delta)]P_f(x) = Y_f(x - \delta)P_f(x) \quad (1.7)$$

B. Middle Ear

The middle ear mechanics are summarized in figure 1.8. The input to the middle ear is sound pressure P_e at the eardrum. The mechanical admittance of the middle ear is defined as:

$$Y_m = [K_m/(i\omega) + R_m + i\omega M_m]^{-1} \quad (1.8)$$

where R_m , M_m , K_m are the damping, mass and stiffness of the middle ear respectively.

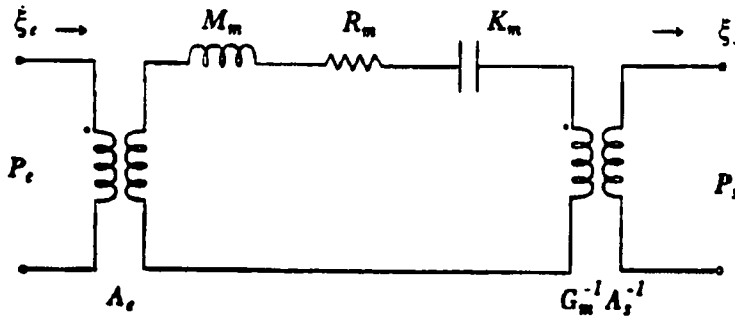


Fig 1.8 Summary of Middle-ear Mechanics [Neely 93]

The middle-ear is coupled to the cochlea by assuming that the pressure at the stapes is $P_s = P_f(0)$ and that the volume velocity at the stapes is

$$\dot{\xi}_s = \dot{\xi}_f(0) \quad (1.9)$$

If we also require that the CP has zero admittance at $x = 0$, then:

$$\xi_f(0) = g_m A_s A_e Y_m P_e - (g_m A_s)^2 Y_m P_f(0) \quad (1.10)$$

where g_m is the 'level gain' of the middle-ear ossicles, A_e is the effective area of the eardrum, and A_s is the area of the stapes footplate.

C. Micro-mechanics

The micro-mechanics models the movement and interaction of the structures within CP. The anatomical structure of a RCS of the cochlear partition is illustrated in figure 1.9. It is assumed that the separation between the BM and the RL will be changed by OHC contraction.

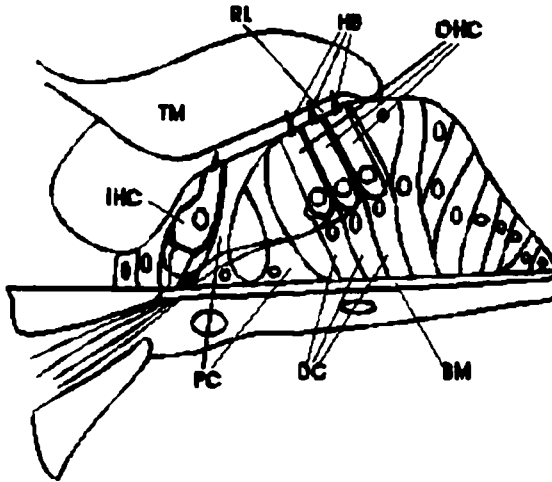


Fig. 1.9 Illustration of the Anatomical Structure of a RCS of the Cochlear Partition

The amount of OHC contraction ξ_c is directly related to lateral deflection ξ_o of the HB at the top of OHC by an OHC gain function H_c , that is:

$$\xi_c = H_c \xi_o \quad (1.11)$$

The effective area of the CP for this section is:

$$A_p = w_b \delta \quad (1.12)$$

where w_b is the effective width of the BM.

The velocity of the BM is directly proportional to the volume velocity of the CP, i.e.

$$\dot{\xi}_b = -A_p \dot{\xi}_p \quad (1.13)$$

The cochlear micromechanics for one RCS are summarized in figure 1.10. The mechanical admittance of BM and TM is:

$$Y_b = [K_b / (i\omega) + R_b + i\omega M_b]^{-1} \quad (1.14)$$

$$Y_t = [K_t / (i\omega) + R_t + i\omega M_t]^{-1} \quad (1.15)$$

where R_b , M_b , K_b are the damping, mass and stiffness of the basilar membrane. R_t , M_t , K_t are the damping, mass and stiffness of the tectorial membrane.

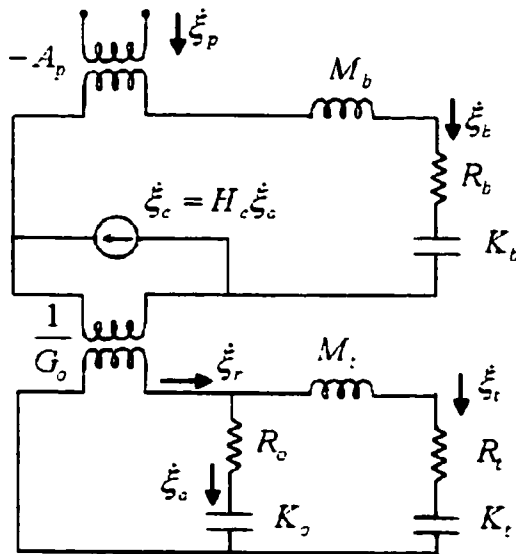


Fig. 1.10 Summary of the Cochlear Micromechanics for One RCS

The mechanical impedance which couples BM and TM is:

$$Z_o = K_o / (i\omega) + R_o \quad (1.16)$$

where K_o , R_o are stiffness of HB of OHC and the viscous damping of the subtektorial fluid.

The displacement ξ_o of the HB is the difference between lateral displacement ξ_r of the RL and lateral displacement ξ_t of the TM,

$$\xi_o = \xi_r - \xi_t \quad (1.17)$$

The lateral RL displacement ξ_r is assumed to be proportional to BM displacement ξ_b minus any contraction ξ_c of the OHC

$$\xi_r = g_o(\xi_b - \xi_c) \quad (1.18)$$

The acoustic admittance of the CP is given by:

$$Y_p = A_p^2 Y_h (1 + g_o H_o Z_o Y_o)^{-1} \quad (1.19)$$

where H_o is the transfer function that relates ξ_o to ξ_b

$$H_o = \xi_o / \xi_b = g_o (1 + g_o H_c + Z_o Y_o)^{-1} \quad (1.20)$$

and H_c is the OHC gain function.

D. Outer Hair Cell

The motility of OHCs is represented in the models by an OHC gain function that describes the ratio of OHC contraction to HB deflection. The OHC gain function is the product of the two separate transducer functions, i.e.:

$$H_c = \gamma T_f T_r \quad (1.21)$$

where γ is a multiplier used to demonstrate the effects of impaired OHC motility.

Mechanoelectric transduction at HB T_f is given by:

$$T_f = g_f / (1 + i\omega\tau_f) \quad (1.22)$$

Electromechanic transduction at the lateral membrane T_r is given by:

$$T_r = g_r / (1 + i\omega\tau_r) \quad (1.23)$$

E. Inner Hair Cell

To calculate the displacement of the HB of the IHC, it is assumed that the HB displacement is proportional to the fluid displacement at high frequencies and proportional to fluid velocity at low frequencies. The transition frequency between these two states will be denoted by f_c . The shearing displacement between TM and RL in the vicinity of the IHCs is less affected (by fraction α) by OHC contraction than by the HB deflection of the OHC. Therefore, the displacement ξ_i of the HB on the IHC is computed as:

$$\xi_i = \{g_o[\xi_o - (1-\alpha)\xi_c] - \xi_r\} [1 + 2\pi f_c / (i\omega)]^{-1} \quad (1.24)$$

The transfer function becomes:

$$H_i = \xi_i / \xi_b = H_o (1 - \alpha g_o H_c) [1 + 2\pi f_c / (i\omega)]^{-1} \quad (1.25)$$

F. Model Results

The tuning curve produced by the example model is shown in figure 1.11. The ordinate indicates the threshold of sound pressure required at the eardrum.

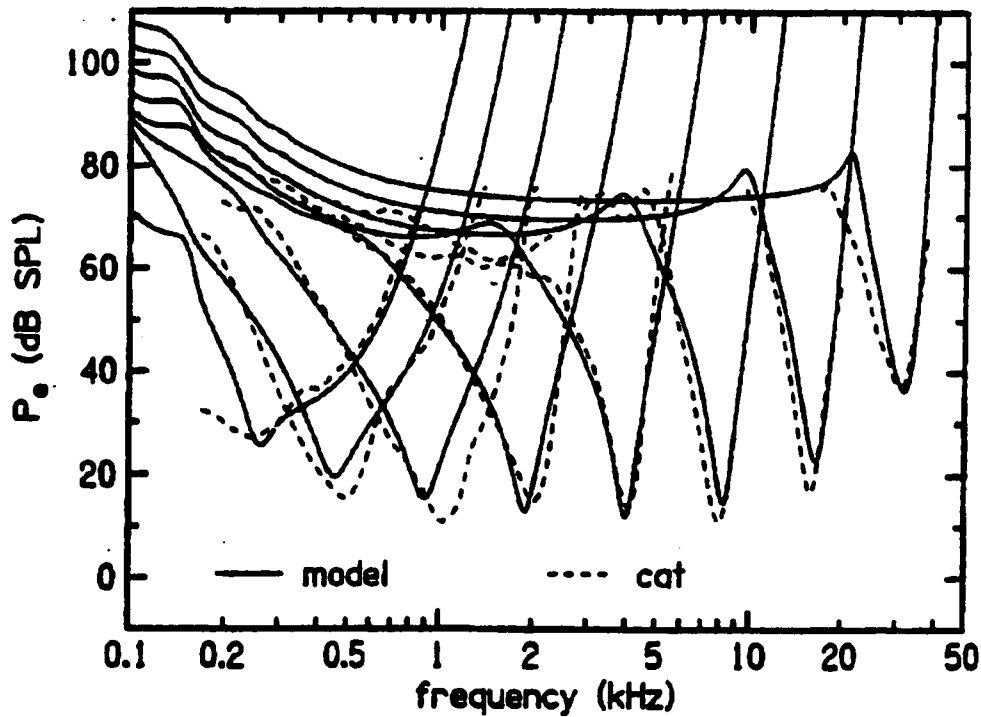


Fig 1.11 The Tuning Curve Produced by the Example Model

It shows good agreement between the model and the measured data. This model is an attractive tool for computational studies of the hearing system, in particular for research involving the interaction of many anatomical components as in studies of the acoustic reflex and of the

cochlear efferent system. It completely models the cochlear mechanisms and accurately simulates the neural tuning curves both on magnitude and on latency.

The main shortcoming of these models is that it needs too much computational burden and memory resource to use on real time speech recognition since it needs too many section (500) to get accurate results.

3.3 Phenomenological Filter Model

Phenomenological filter models are designed just to account for observed neural tuning phenomena, without detailed concern for biological mechanisms. Examples of these models are Patterson's model [Patterson, Holdsworth, and Allerhand 92], multiple band-pass non-linearity (MBPNL) model [Goldstein 95], and dual resonance non-linearity (DRNL) model [O'Mard, Lopez-Poveda, and Meddis].

The block diagram of DRNL is shown in figure 1.12. The combined effect of the outer-/middle ear transfer function can be modeled as a simple band-pass filter. The DRNL filter contains two parallel filter paths, one more sensitive and sharply tuned, and the other less sensitive and broadly tuned. Both paths receive the input signal and the resulting output is the summation of the outputs of the two paths. The narrow filter path contains the compressive memory-less non-linearity between two filters. The non-linearity compresses only the part of a signal that exceeds the threshold amplitude. The low-pass filter serves two purposes: it increases the high frequency slope of the filter and also introduces a phase change. The wide filter path is attenuated to reduce its sensitivity, relative to the narrow filter path.

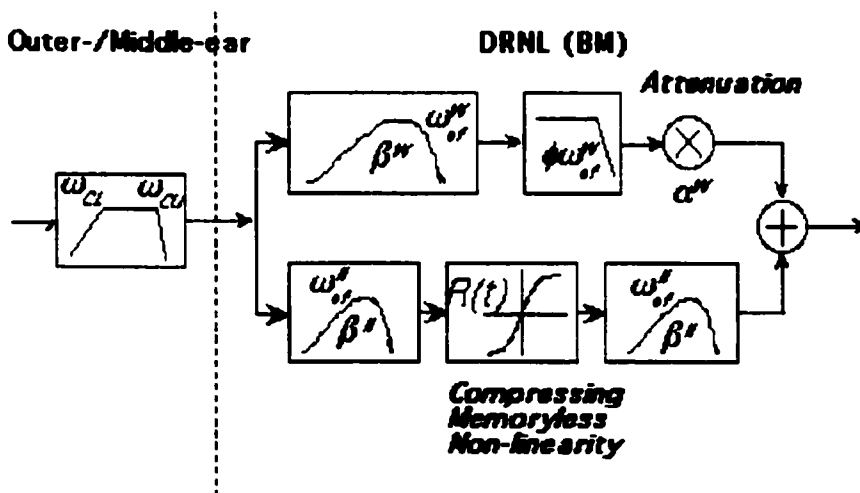


Fig. 1.12 The Block Diagram of DRNL

Figure 1.13 shows the filter-bank model tuning curve with CF = 8 kHz (solid line with circles) compared with BM data. DRNL model closely simulates experimental results obtained from mechanical cochlear experiments. This purely passive model demonstrates a center frequency shift, and filter width increase with a rise in stimulus level. One consequence of this robust, passive structure is that it allows a great deal of freedom in the choice of parameters. Since it entirely consists of the gammatone filter and Butterworth filter it represents it is a causal system. And it gives a transfer function with the amplitude and phase characteristics of a realizable physical filter. Moreover, the filter is minimum phase. The output of a bank of DRNL can be calculated by an efficient computer algorithm. The problem in this model is that it is not accurate enough to simulate the neural tuning curve obtained from the experiments.

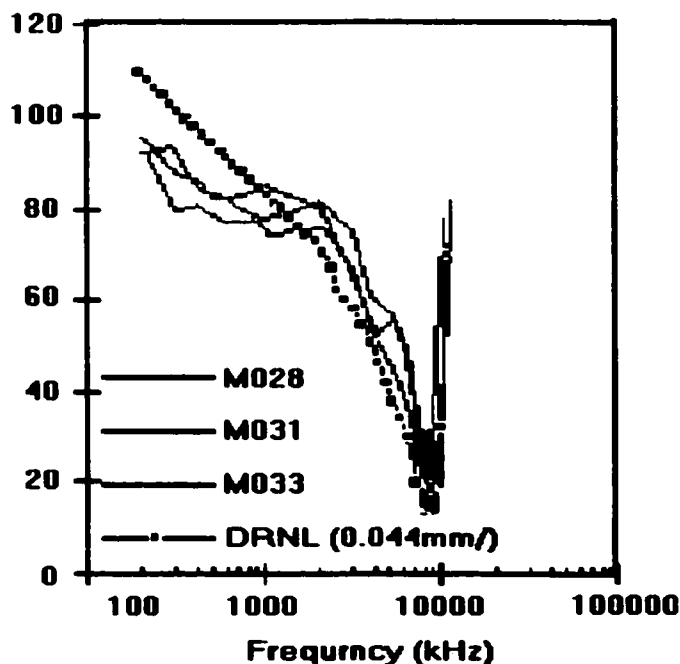


Fig. 1.13 The Filter-bank Model Tuning Curve with CF = 8 kHz (solid line with circles)
Compared with BM Data.

3.4 The Problems in the Existing Cochlear Model and Objective of Our Research

The problems in the existing cochlear model are:

1. Low-pass filter sections cascade model and phenomenological filter model are not enough accurate to meet the requirement of speech recognition.
2. Transmission-line model needs too much computational burden to meet the real-time speech recognition.

Our research objective is to design more accurate and more efficient cochlear model to meet the requirement of real-time speech recognition.

4. The Genetic Search Approach

Genetic algorithms are inspired by Darwin's theory about evolution. The underlying principles of GAs were first published by Holland [Holland 62] and extensively explored by Goldberg [Goldberg 89]; [De Jong 88]. Recently, genetic algorithms (GAs) have been applied to IIR refinement problems. In these approaches the simple and basic GA forms are introduced for multimodal optimization in IIR filter designs.

4.1 The Outline of the GA

The outline of the basic genetic algorithm is summarized in table 1.1. The algorithm is started with a *set of solutions* (represented by *chromosomes*) called *population*. Solutions from one population are taken and used to form a new population. This is motivated by a hope, that the new population will be better than the old one. Solutions that are selected to form new solutions (*offspring*) are selected according to their fitness - the more suitable they are the more chances they have to reproduce. This is repeated until some condition (for example number of populations or improvement of the best solution) is satisfied.

Table 1.1 The Outline of Basic Genetic Algorithm

Step 0:	Initialization: Generate random population of n chromosomes (suitable solutions for the problem).
Step 1:	Fitness evaluation: Evaluate the fitness $f(x)$ of each chromosome x in the population .
Step 2:	New population creation: Create a new population by repeating following steps until the new population is complete.
Step 2.0:	Reproduction: Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected).
Step 2.1:	Crossover: With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.
Step 2.2:	Mutation: With a mutation probability mutate new offspring at each locus (position in chromosome).
Step 2.3:	Accepting: Place new offspring in a new population.
Step 3:	Replace: Use new generated population for a further run of algorithm.
Step 4:	Test: If the end condition is satisfied, <i>stop</i> , and return the best solution in current population.
Step 5:	Loop: Go to step 2.

4.2 Reproduction

Reproduction is a process in which individual chromosomes are copied according to their fitness value. According to Darwin's evolution theory the better ones should survive and create more children. A fitness $f(i)$ is assigned to each individual in the population, where high numbers denote good fit. The reproduction (parents selecting) process is conducted by spinning simulated biased roulette wheel whose slots have different sizes proportional to the fitness values of the individuals. This technique is called *roulette-wheel parent selection*. It can be simulated by the algorithm in table 1.2.

Table 1.2 The Simulation Algorithm for Roulette-wheel Parent Selection

Step 0:	Sum: Calculate sum of all chromosome fitnesses in population - sum S .
Step 1:	Reproduce: Generate random number from interval $(0, S) - r$.
Step 2:	Loop: Go through the population and sum fitnesses from 0 to sum s . When the sum s is greater than or equal to r , stop and return the chromosome where you are.

4.3 Elitism

When creating new population by crossover and mutation, we have a big chance, that we will lose the best chromosome. Elitism will first copy the best chromosome (or a few best chromosomes) to new population. The rest is done in classical way. Elitism can very rapidly increase performance of GA, because it prevents losing the best found solution.

4.4 Crossover

Reproduction directs the search towards the best existing individuals but does not create any new individuals. In nature, an offspring has two parents and inherits genes from both. The main operator working on the parents is crossover, which happens for a selected pair with a crossover probability p_c . Crossover selects genes from parent chromosomes and creates a new offspring. The simplest way how to do this is to choose randomly some crossover point and everything before this point copy from a first parent and then everything after the crossover point copy from the second parent. Crossover can then look like this (| is the crossover point):

Chromosome 1	10001 00100110110
Chromosome 2	11011 11000011110
Offspring 1	10001 11000011110
Offspring 2	11011 00100110110

4.5 Mutation

Although reproduction and crossover produce many new chromosomes, they do not introduce any new information into the population at the bit level. This will fall all solutions in population into a local optimum of solved problem. As a source of new bits, mutation is introduced and is applied with a low probability p_m . Mutation changes randomly the new offspring. For binary encoding we can switch a few randomly chosen bits from 1 to 0 or from 0 to 1. Mutation can then be following:

Chromosome 1	1101111000011110
Chromosome 2	1101100100110110
Offspring 1	1100111000011110
Offspring 2	1101101100110100

4.6 GA for IIR Filter Optimization

As a randomization process the GA is used to solve IIR coefficient optimization problem. The algorithm is summarized in table 1.3 [Arsian and Horrocks 95], [Haseyama 96], [Kosir and Tasic 95].

Table 1.3 Genetic Search Learning Algorithm for IIR Filter Design

Step 0:	Initialization: Given a fixed IIR structure, initialize the filter coefficients (the initial chromosome) by some small real random numbers.
Step 1:	Fitness evaluation: Calculate the fitness for each set of coefficients. If the stopping criterion is satisfied, STOP; Otherwise, do following steps.
Step 2:	Elitism: Copy a few best sets of coefficients into the new population. Delete a few worst sets of coefficients from the population.
Step 3:	Reproduction: Choose the sets of coefficients to new population according their fitness. The more the fitness it has the more chance it will be selected.
Step 4:	Crossover: Randomly match the sets of the coefficients with the crossover probability and apply the crossover operator at the random position in the coefficient sets.
Step 5:	Mutation: Randomly add a small real number to the current filter coefficient set at some probability.
Step 6:	Replace: construct the new population by the new sets of coefficients that are created from step 2 to step 5
Step 7:	Loop to step 1

In this approach, the set of IIR filter coefficients is first encoded into a chromosome as a list of real numbers. The GA operators, such as elitism, reproduction, crossover, mutation, are used to create new solution. Whenever the stopping criterion is met the GA search will be stopped. The mutation is applied by randomly perturbing the current filter coefficient set at some probability. The coefficients are produced in a certain controlled manner; particularly, they need to satisfy stability condition for the filter. Among these new sets of coefficients the chromosome with the best fitness will be selected as survivor.

5 Lattice Form Realization

5.1 Problems in the Traditional Canonic Direct-form

The implementation of the filter has a great impact on the actual performance of the filter. The main problems in the traditional canonic direct-form is that the trade-off between overflow and round-off noise and finding the optimal point between overflow and round-off noise. To avoid the overflow we need to minimize the input signal level. But to increase SNR we need to increase the input signal level [Massie 93]. The standard second-order transfer function is given as:

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \quad (1.26)$$

For direct form realization as Q increases or as the bandwidth decreases, the round-off noise of the direct-form filter will increase without bound. In general the range of coefficients in the direct form is ± 1 for a_2 , and ± 2 for a_1 and b_1 , it could be inconvenient to implement coefficient multiplication by a number greater than 1 in many fix-point architectures. The internal state variables in the direct form can have a very large magnitude, depending on the pole and zero locations. This requires either extra headroom (extra bits) to allow for word growth in internal registers, or scaling the input signal to avoid internal overflow. The z^{-1} terms ($-2r\cos\theta$) indicate that the center frequency interacts with Q in this structure. The z^{-2} terms (r^2) determine the radius of the poles and zeros. While the radius (r) is changed the z^{-1} terms must be changed, or the frequency (θ) will be changed too. It is desirable to have the center frequency and Q vary independently to make the calculation of the filter coefficients easier. Moreover, it is quit possible to create unstable filter with the direct-form implementation.

5.2 Principles of the Four-multiply Normalized Ladder Filter

The basic principle of the four-multiply, normalized ladder filter is that any stable rational transfer function can be decomposed into all-pass subsystems with one additional output multiply, which is required for L_2 re-normalization to restore the defined output norm. If the all-pass subsystem is implemented using ladder or lattice forms then the coefficients for the three parameters (peak gain, bandwidth, and center frequency) in the parametric second section are independent variables. If the all-pass subsystem is implemented using a four-multiply normalized ladder form, then each internal variable of the all-pass subsystem is automatically scaled in the L_2 sense.

5.3 Regalia and Mitra's Method

Figure 1.14 shows implementation diagrams of lattice form using Regalia and Mitra's method.

The overall transfer function is given by:

$$F(z) = \frac{1}{2} [1 + A(z)] + \frac{1}{2} G [1 - A(z)] \quad (1.27)$$

where G controls peak gain and $A(z)$ is an all-pass subsystem.

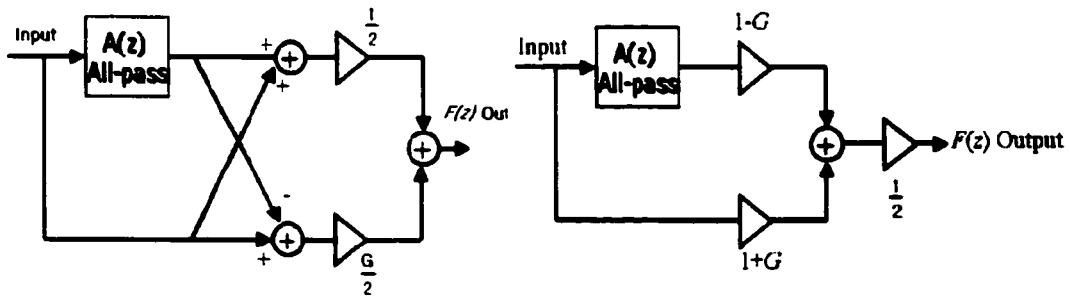


Fig 1.14 Diagrams of Lattice Form Using Regalia and Mitra's Method.

5.4 First-order Lattice Filter

Figure 1.15 shows the diagram of implementation of the first-order all-pass filter. The the first-order all-pass filter is:

$$A(z) = \frac{k_2 + z^{-1}}{1 + k_2 z^{-1}} \quad (1.28)$$

where:

$$k_2 = \frac{\tan(\Omega/2) - 1}{\tan(\Omega/2) + 1} \quad (1.29)$$

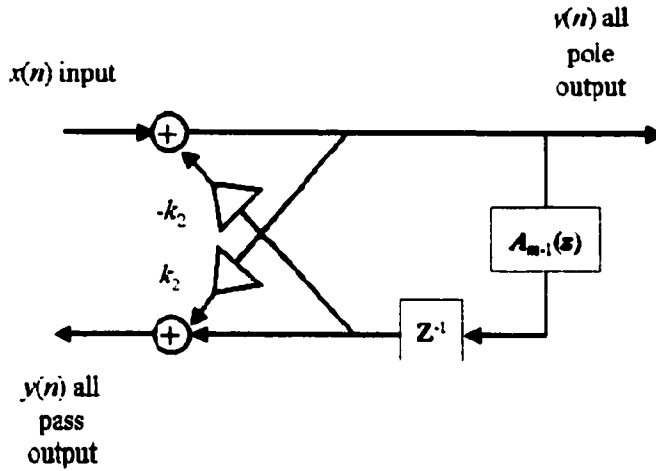


Fig 1.15 Diagram of Implementation of the First-order All-pass Filter

Including the nested element, the first-order all-pass filter is given as:

$$A(z) = \frac{k_2 + z^{-1} A_{m-1}(z)}{1 + k_2 z^{-1} A_{m-1}(z)} \quad (1.30)$$

5.5 Second-order Lattice Filter

Figure 1.16 shows the implementation diagram of second-order all-pass filter. Its all-pass subsystem $A(z)$ is given by:

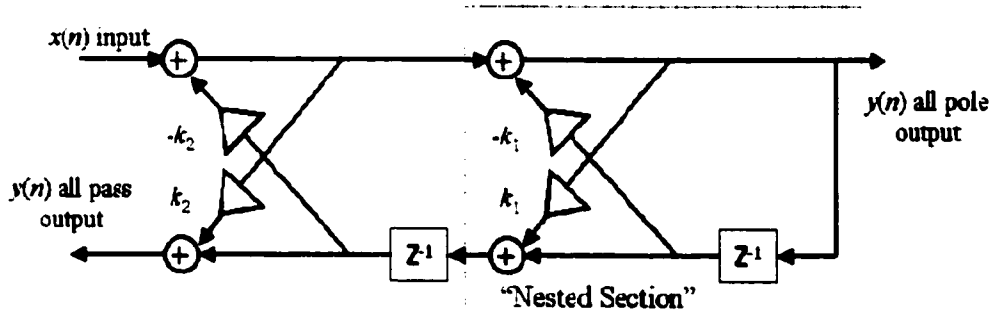
$$A(z) = \frac{z^{-2} + k_1(1 + k_2)z^{-1} + k_2}{1 + k_1(1 + k_2)z^{-1} + k_2z^{-2}} \quad (1.31)$$

Its parameter design equations are:

$$k_1 = -\cos \frac{2\pi f}{f_s} \quad (1.32a)$$

$$k_2 = r^2 = \frac{1 - \tan(\Omega/2)}{1 + \tan(\Omega/2)} \quad (1.32b)$$

where f_s is the sampling frequency and f is the filter center frequency. r is the pole radius and Ω the bandwidth of the filter.



where: $k_1 = -\cos \omega_0$

$$A(z) = \frac{k_2 + k_1(1 + k_2)z^{-1} + z^{-2}}{1 + k_1(1 + k_2)z^{-1} + k_2z^{-2}}$$

$$k_2 = \frac{1 - \tan(\Omega/2)}{1 + \tan(\Omega/2)}$$

$$G = \frac{(m_0/\Omega)}{\sqrt{1 - 1/4(\omega_0/\Omega)^2}}$$

Fig. 1.16 Implementation Diagram of Second-order All-pass Filter

5.6 The Four-multiply Markel-Gary Normalized Ladder

In four-multiply Markel-Gary normalized ladder realization every node is normalized in the L_2 sense. This can potentially be very powerful feature, especially in the time-varying case. No matter what the position of the poles and zeros, the power at each node is unity for a white-noise input. The noise amplitude does not vary as a function of the pole-zero position (i.e. bandwidth or frequency). It is a constant and dependent only on the filter order (at all-pass output). It makes the resonant frequency and Q controlled by independent parameters. There is no overflow limit cycle in normalized ladder filters. As long as the k coefficients are less than 1, the filter would be stable. When the coefficients for a stable normalized ladder filter are time varying, the filter can be proved to still be stable. The single disadvantage of the normalized ladder is a number of multiplies needed. The total cost is about four times the cost of a direct-form structure for the same order filter.

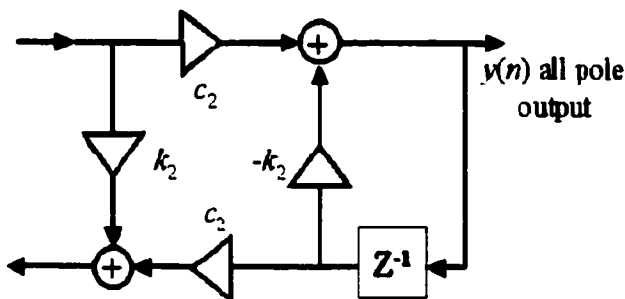


Fig. 1.17 The Structure of All-pass Subsection of The First Order Normalized Ladder Filter

Figure 1.17 shows the structure of all-pass subsection of the first order normalized ladder filter.

Its transfer function is given by:

$$A(z) = \frac{k_2 + z^{-1}}{1 + k_2 z^{-1}} \quad (1.34a)$$

where:

$$k_2 = \frac{\tan(\Omega/2) - 1}{\tan(\Omega/2) + 1} \quad (1.34b)$$

Ω is the low-pass or high-pass filter cut-off frequency.

Figure 1.18 shows the structure of all-pass subsection of the second order normalized ladder filter. Its transfer function is given by:

$$A(z) = \frac{k_2 + k_1(1 + k_2)z^{-1} + z^{-2}}{1 + k_1(1 + k_2)z^{-1} + k_2 z^{-2}} \quad (1.35a)$$

where:

$$k_1 = -\cos \omega_0 \quad (\omega_0 \text{ center frequency}) \quad (1.35b)$$

$$k_2 = \frac{1 - \tan(\Omega/2)}{1 + \tan(\Omega/2)} \quad (\Omega \text{ bandwidth}) \quad (1.35c)$$

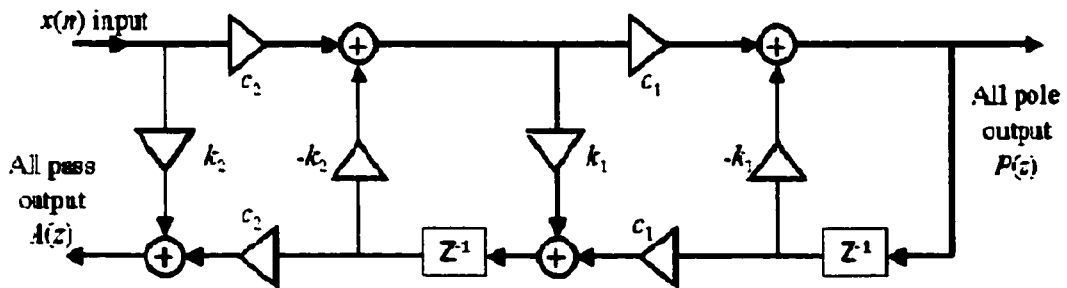


Fig. 1.18 Structure of All-pass Subsection of the Second Order Normalized Ladder Filter

6. The Ensemble Interval Histogram (EIH) Representation

Ensemble interval histogram (EIH) model comprises two stages, one that models the pre-auditory-nerve section and another stands for the post auditory-nerve section of the auditory periphery. The EIH model is schematically illustrated in figure 1.19 [Ghitza 92]. The cochlear filters are used to model mechanical motion of the basilar membrane, which is reviewed in section 3 of this chapter.

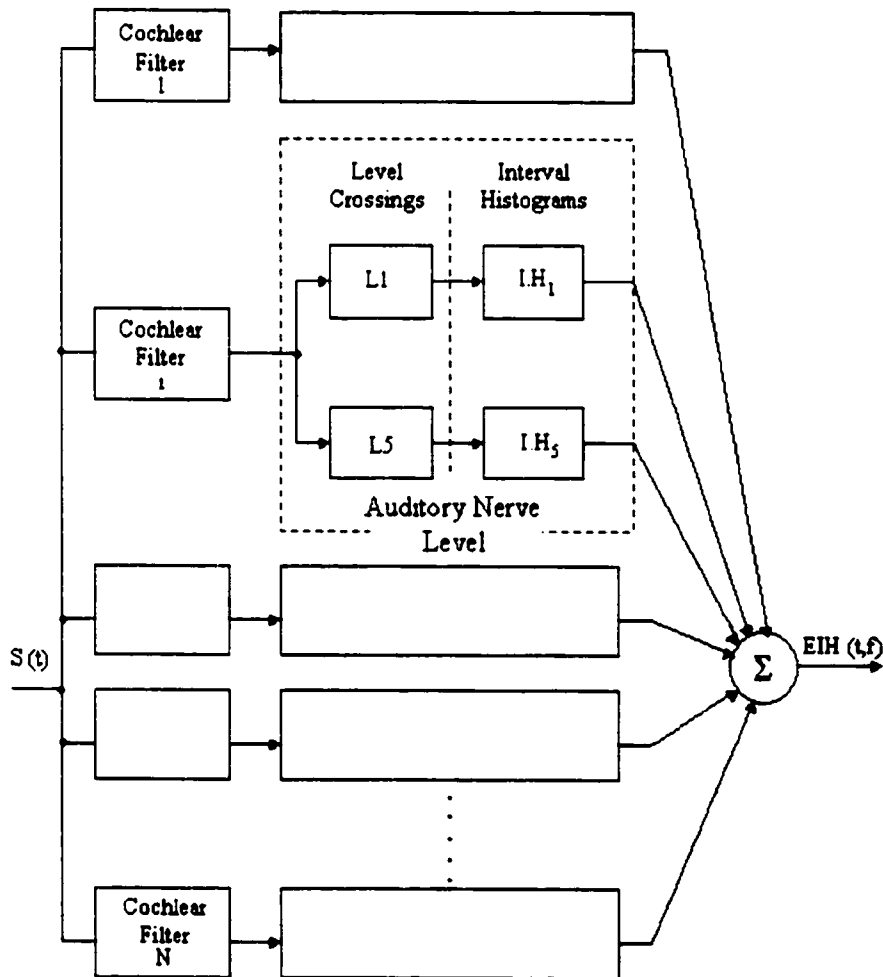


Fig 1.19 Block Diagram of EIH

6.1 Level-crossing

The auditory nerve activity is simulated with an array of level-crossing detectors at the output of each cochlear filter. A neural firing is simulated as the positive-going level crossing. The thresholds are distributed across a range of positive levels, to account for the half-wave rectification nature of the IHC receptor potential. The threshold value assigned to the level j of every filter is a random Gaussian variable, with a mean L_j , and a standard deviation $s_j=0.2L_j$.

6.2 Measurement of Synchrony and Instantaneous Rate Across the Simulated Fibers

An estimate of the interval probability density function of a given level can be obtained by computing a histogram of the intervals from the point process data produced by the level-crossing detector. The similarity between all individual interval probability density functions is measured by collecting the individual histograms into one ensemble interval histogram (EIH). In order to obtain an auditory representation in the frequency domain the histogram of the inverse intervals is computed, i.e., distributing the reciprocal of the intervals in a histogram.

C h a p t e r 2

Cochlear Filter Model Description

Section 1 Introduction

Section 2 Outer/middle Ear Combined Bandpass Filter

Section 3 Filterbank

Section 4 Second Filter

Section 5 Normalized Lattice Form Realization of 2nd Order Digital Filter

Section 6 Filter Optimization Using GA

Section 7 Half Wave Rectification

Section 8 Nonlinear Stages

Section 9 Summary

1. Introduction

The model outlined in figure 2.1 consists of a number of stages: (1) outer/middle combined bandpass filter; (2) filterbank; (3) second filter; (4) half-wave rectification; (5) non-linearity.

Each of these stages will be described individually.

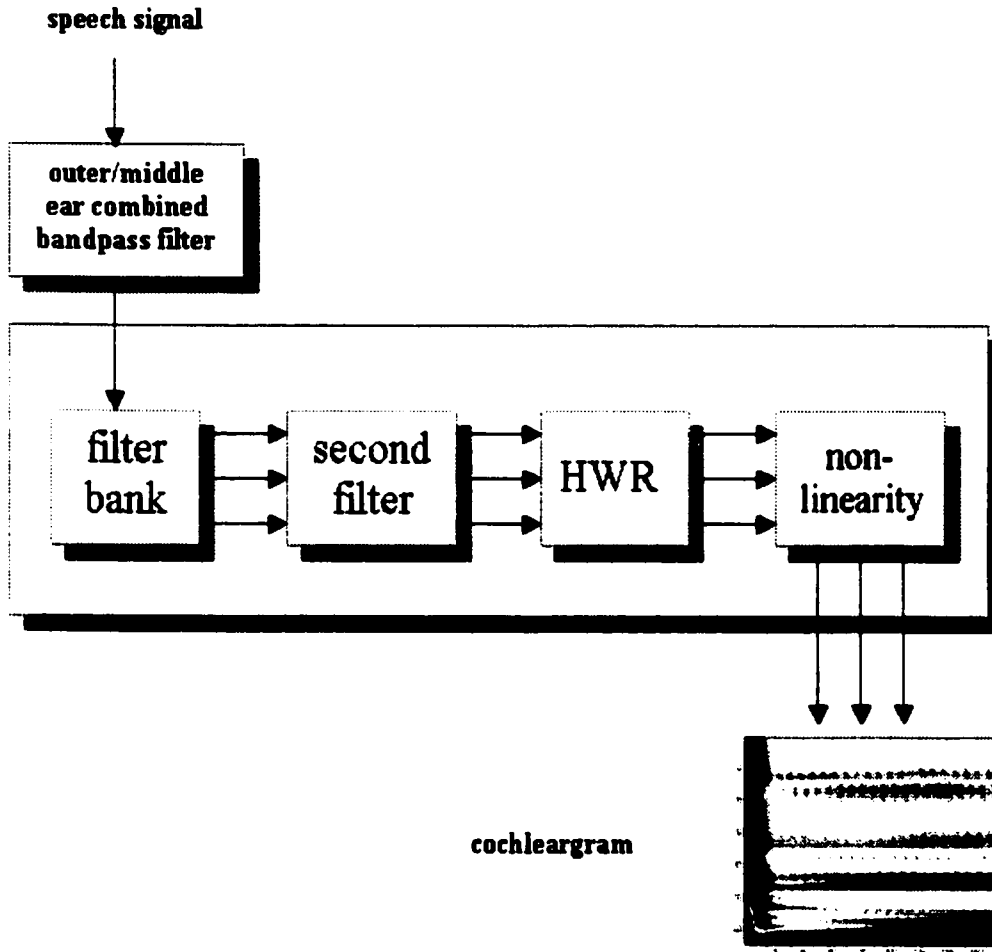


Fig 2.1 Block Diagram of the Cascaded Low-pass Cochlear Model

2. Outer/middle Ear Combined Bandpass Filter

Sound entering the outer ear is subject to a pressure gain at the tympanic membrane relative to the ear canal; this pressure gain is maximal in the region between 2 and 5 kHz [Nedzelnsky 74].

The middle ear, which couples sound energy from the external auditory meatus to the cochlea also has a bandpass pressure-transfer function but, on this occasion, the peak is near 1 kHz and has a much steeper slope at the low frequencies. These two functions have been combined to obtain an approximately flat-topped bandpass function.

In the model the outer/middle ear is modeled by a highpass filter with a cutoff frequency at 1000 Hz.

Letting f_0 denote the cutoff frequency of the highpass filter and f_s denote sample frequency. The highpass filter is given by:

$$H(z) = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1}} \quad (2.1a)$$

where:

$$b_0 = \frac{2}{\tan(\Omega/2) + 1} \quad (2.1b)$$

$$b_1 = -\frac{2}{\tan(\Omega/2) + 1} \quad (2.1c)$$

$$a_1 = \frac{\tan(\Omega/2) - 1}{\tan(\Omega/2) + 1} \quad (2.1d)$$

and

$$\Omega = \frac{2\pi f_0}{f_s} \quad (2.1e)$$

A zero at the Nyquist rate ($1+z^{-1}$) is added and it is combined with highpass filter to form a bandpass filter.

In order to convert pressure waves into basilar membrane motion a differentiator (a term of $1-z^{-1}$) is followed [Slaney 88].

Meanwhile, the gain of the filter is set so that it has unity gain at $f_s/4$. Figure 2.2 presents the frequency response of the outer/middle ear.

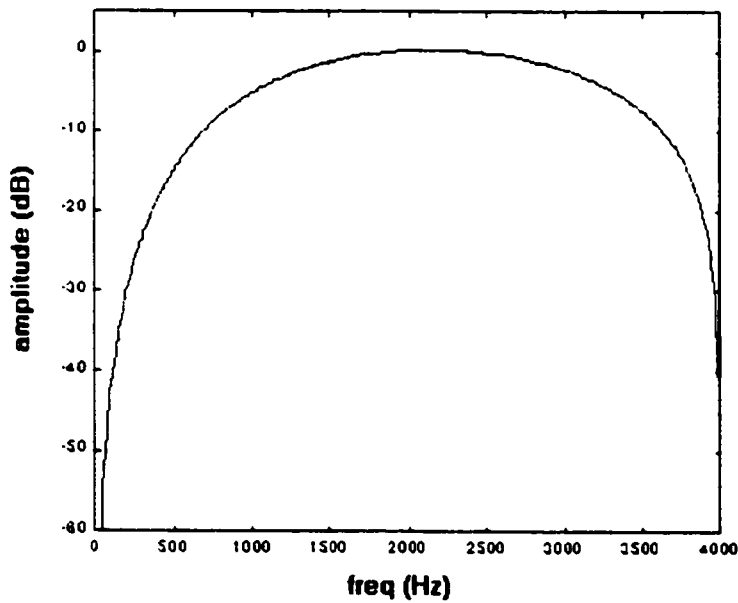


Fig 2.2 Frequency Response of the Outer/middle Ear Combination

3. Filterbank

The behavior of the bandpass filtering effect of the basilar membrane was achieved using a set of cascaded digital IIR filter [Lyon 82]. The net filtering effect at a given location is given by:

$$G_k(z) = \prod_{i=1}^k H_i(z) \quad (2.2)$$

where $H_i(z)$ is the transfer function of each filter section in the cascade.

Each section in the cascaded lowpass filter consists of a notch and a resonator filter. As combination of them each lowpass section consists of two zeroes and two poles as given below:

$$H_i(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \quad (2.3)$$

In next few sections the selection of the parameters of these lowpass filters will be introduced individually [Lyon82].

3.1 Definitions

Before designing the filter section some parameter definition will be given first.

A. EarQ

EarQ defines the Q value of the ear. It is used to calculate the bandwidth of the filter. In this model the default value is 8.

B. EarBreakFreq

This constant is used to calculate the bandwidth of the lower center frequency filters. It takes 1000 in this model.

C. StepFactor

StepFactor is the overlap fraction of the filter bandwidth, e.g. *StepFactor* 0.25 means in any filter bandwidth there are 4 filter overlapped.

D. EarZeroOffset

EarZeroOffset determines the offset between the center frequency and the zero.

E. EarSharpness

EarSharpness sets how much sharper the zero is than its associated pole.

3.2 Bandwidth

The bandwidth of each ear filter is related to the center frequency of filter directly as given :

$$Bandwidth(CF) = \sqrt{\frac{CF^2 + EarBreakFreq^2}{EarQ}} \quad (2.4)$$

Figure 2.3 demonstrates the relationship between bandwidth and filter center frequency (sample frequency 8000 Hz).

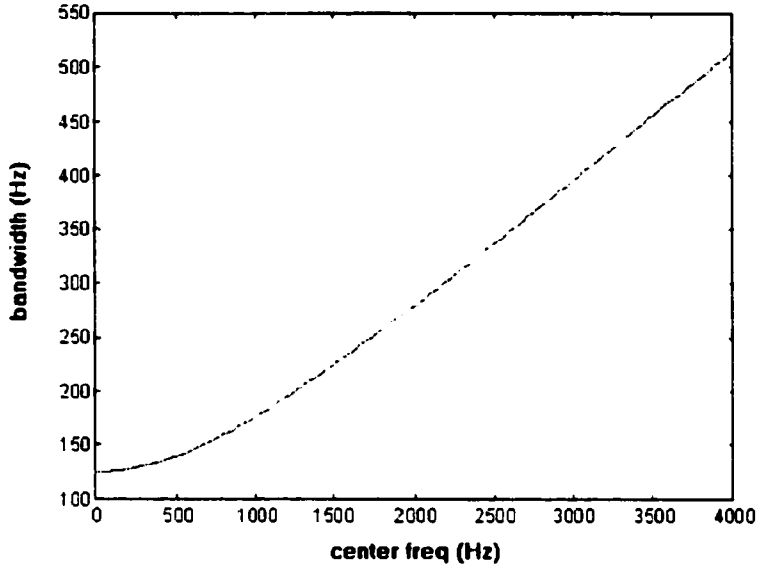


Fig 2.3 Channel Bandwidth vs. Channel Center Frequency

3.3 Center Frequency

First, the lowest and the highest center frequencies are given as:

$$HighestCF = \frac{f_s}{2} - Bandwidth\left(\frac{f_s}{2}\right) \bullet StepFactor \bullet (EarZeroOffset - 1) \quad (2.5)$$

$$LowestCF = \frac{EarBreakFreq}{\sqrt{4EarQ^2 - 1}} \quad (2.6)$$

For i th filter section we have:

$$CF(i) = CF(i - 1) - Bandwidth(i - 1) \bullet StepFactor \quad (2.7)$$

So each filter center frequency can be calculated using dynamic programming from the highest center frequency.

The associated zero is given by:

$$\text{ZeroCF}(i) = \text{CF}(i) + \text{Bandwith}(i) \bullet \text{StepFactor} \bullet \text{EarZeroFactor} \quad (2.8)$$

Figure 2.4 lists each resonant and notch center frequency in the model (sample rate 8000Hz).

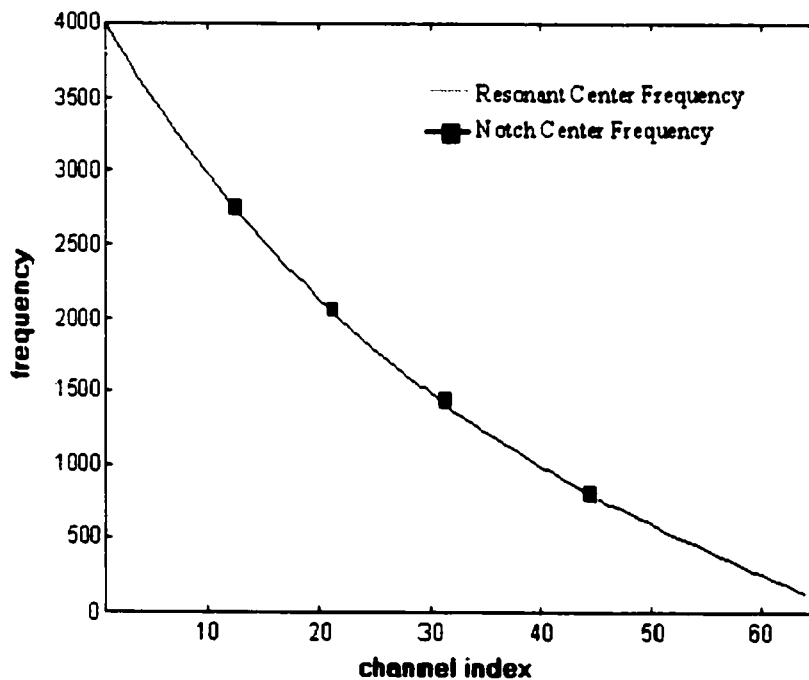


Fig. 2.4 Center Frequencies of Resonant and Notch vs. Channel Index

3.4 Filter Q Values

According to the definition of Q value in second order filter the filter Q value is determined by its filter center frequency and bandwidth. For i th filter section in cascade its zero Q and pole Q is given as:

$$ZeroQ(i) = \frac{ZeroCF(i) \bullet EarSharpness}{Bandwidth(i)} \quad (2.9)$$

$$PoleQ(i) = \frac{CF(i)}{Bandwidth(i)} \quad (2.10)$$

Figure 2.5 shows the relationship between Q values of pole and zero in the model

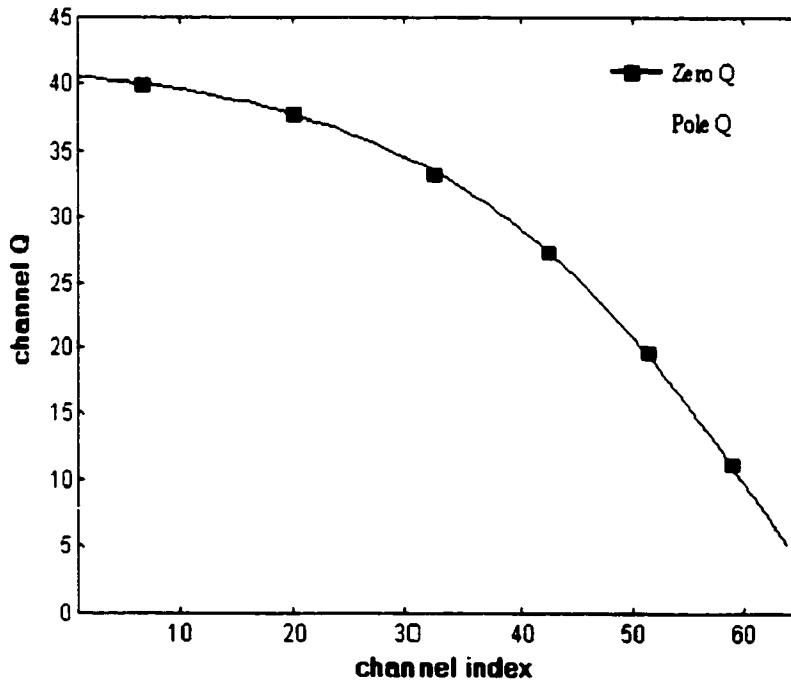


Fig. 2.5 Q Values of Pole and Zero vs. Channel Index

3.5 Filter Gain

Because of each filter channel is provided a differentiator and the gain of an ideal differentiator is proportional to the frequency it makes the lower center frequency channels have lower output.

The gain of the filter should be adjusted so that differentiator

has the unity gain at the center frequency for each channel. The gain of the i th filter section is given as:

$$Gain(i) = \frac{CF(i-1)}{CF(i)} \quad (2.11)$$

where: $Gain(1) = 1$

Figure 2.6 shows the zero and pole distribution on z plane of the second order filters in the cascaded model.

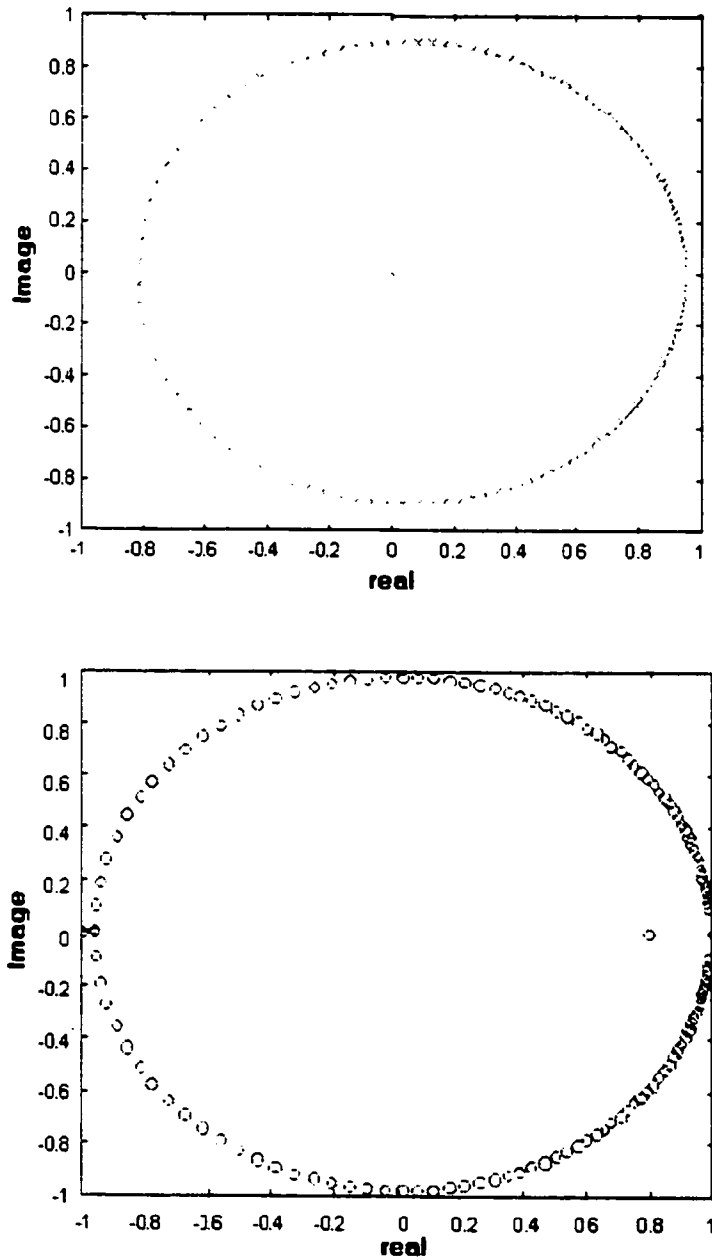


Fig. 2.6 63 Pairs of Poles (Upper) and Zeros (Lower) Distribution on z Plane

4. Second Filter

Tuning-curve measurements, such as those of Kiang and Moxon [Kiang and Moxon 74] show a notch in the frequency response approximately one octave below the pole center frequency. This effect can be modeled by adding a second filter to the filter section between the bandpass filter and the half wave rectification [Kates 93].

For i th filter section the center frequency of the second filter is given as:

$$SecondZeroCF(i) = \frac{CF(i)}{2} \quad (2.12)$$

$$SecondPoleCF(i) = CF(i) \quad (2.13)$$

and the Q value of the second filter is given as:

$$SecondPoleQ(i) = 1.5(1 + CF(i)) \quad (CF(i) \text{ in kHz}) \quad (2.14)$$

$$SecondZeroQ(i) = SecondPoleQ(i) \times 2 \quad (2.15)$$

5. Normalized Lattice Form Realization of 2nd Order Digital Filter

Figure 2.7 shows the realization of the second order filter using normalized lattice form [Massie 93]. This method implements the zeros directly on the top of the Markel-Gary normalized ladder through selecting coefficients vector v .

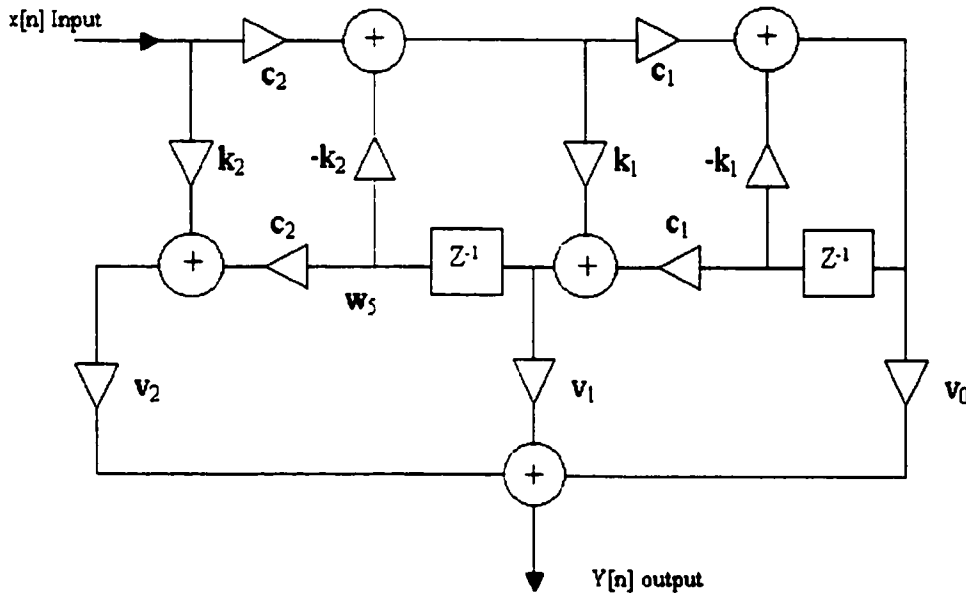


Fig. 2.7 Diagram of Normalized Lattice Form Realization

Assuming we design a second order filter with sample rate at f_s , resonant frequency at f_p , notch frequency at f_z , pole Q value Q_p , and zero Q value Q_z . Let:

$$f_{zn} = f_z / f_s \quad (2.16)$$

$$f_{pn} = f_p / f_s \quad (2.17)$$

$$r_z = e^{\frac{\pi f_{zn}}{Q_z}} \quad (2.18)$$

$$r_p = e^{\frac{\pi f_{pn}}{Q_p}} \quad (2.19)$$

$$\theta_z = 2\pi f_{zn} \cdot \sqrt{1 - \frac{1}{4Q_z^2}} \quad (2.20)$$

$$\theta_p = 2\pi f_{pn} \cdot \sqrt{1 - \frac{1}{4Q_p^2}} \quad (2.21)$$

Then each coefficient in the diagram is given as following:

$$k_2 = r_p^2 \quad (2.22)$$

$$k_1 = \frac{-2r_p \cos(\theta_p)}{1 + k_2} \quad (2.23)$$

$$c_1 = \sqrt{1 - k_1^2} \quad (2.24)$$

$$c_2 = \sqrt{1 - k_2^2} \quad (2.25)$$

$$v_2 = r_z^2 \quad (2.26)$$

$$v_1 = -2r_z \cos(\theta_z) - \frac{v_2 \cdot k_1 (k_2 + 1)}{c_2} \quad (2.27)$$

$$v_0 = \frac{1 - v_1 \cdot c_2 \cdot k_1 - v_2 \cdot k_2}{c_1 \cdot c_2} \quad (2.28)$$

6. Filter Optimization Using GA

Table 2.1 lists the genetic algorithm for the filter parameter optimization.

Table 2.1 The Diagram of Filter Optimization Using GA

Inputs	Initial coefficients: <i>initial</i>
	Target filter response curve: <i>target</i>
	Precision: <i>d</i>
	Previous filter response: <i>RespPre</i>
Step 0	Randomly create 11×10 ((number of new population-1)*number of coefficients) real number between -0.01 to 0.01 and add them with the initial coefficients to create 10 new sets of coefficients. These 11 sets and the initial set consist of first run population.
Step 1	Calculate the <i>fitness</i> for each set of coefficients by: $fitness = \frac{1}{\sum \frac{(resp - target)^2}{resp^2}}$
Step 2	If $fitness < d$ and number of runs < 5000 do following steps, else save and print the most fit set of coefficient
Step 3	Add the most fit 2 sets directly to the new generation and eliminate the least fit 2 sets
Step 4	Randomly match the existing sets and randomly select the crossover points to process the crossover
Step 5	Use mutation rate as 0.1 and randomly select the mutation points, then add a real number (-0.001 to 0.001) to this coefficient.
Step 6	Calculate the fitness for each set of coefficient and get maximum fitness.

7. Half Wave Rectification

The outputs of second filters are passed to a bank of half wave rectifications which model the directional behavior of the inner hair cells, and cut the energy of the signal by approximately two [Lyon 82]. The outputs of the HWRs can be used to represent neural response. HWRs have the function to drop the negative portions of the waveform, i.e.:

$$y(n) = \max(0, x(n)) \quad (2.29)$$

where $x(n)$ is the input of the HWR at time n and $y(n)$ is its output of the HWR.

Figure 2.8 shows the relationship of the input and output of HWR.

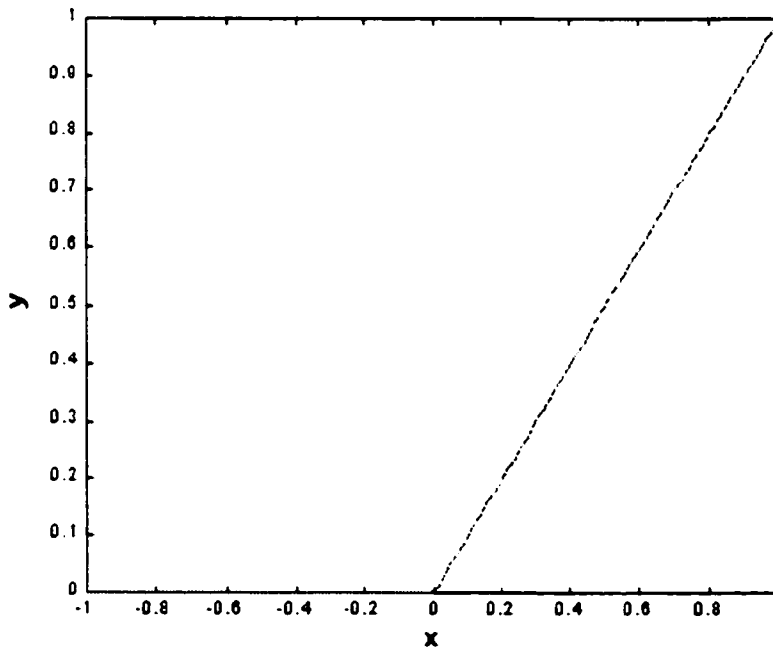


Fig. 2.8 Relationship of the Input and Output of HWR

8. Nonlinear Stages

The nonlinear stages in the model represent the adaptive abilities in the auditory system [Lyon 82]. In this model each channel is followed by four cascaded automatic gain control (AGC) stages with different time constants and different target values that constitute of the nonlinear stage. Figure 2.9 shows the realization of a single AGC.

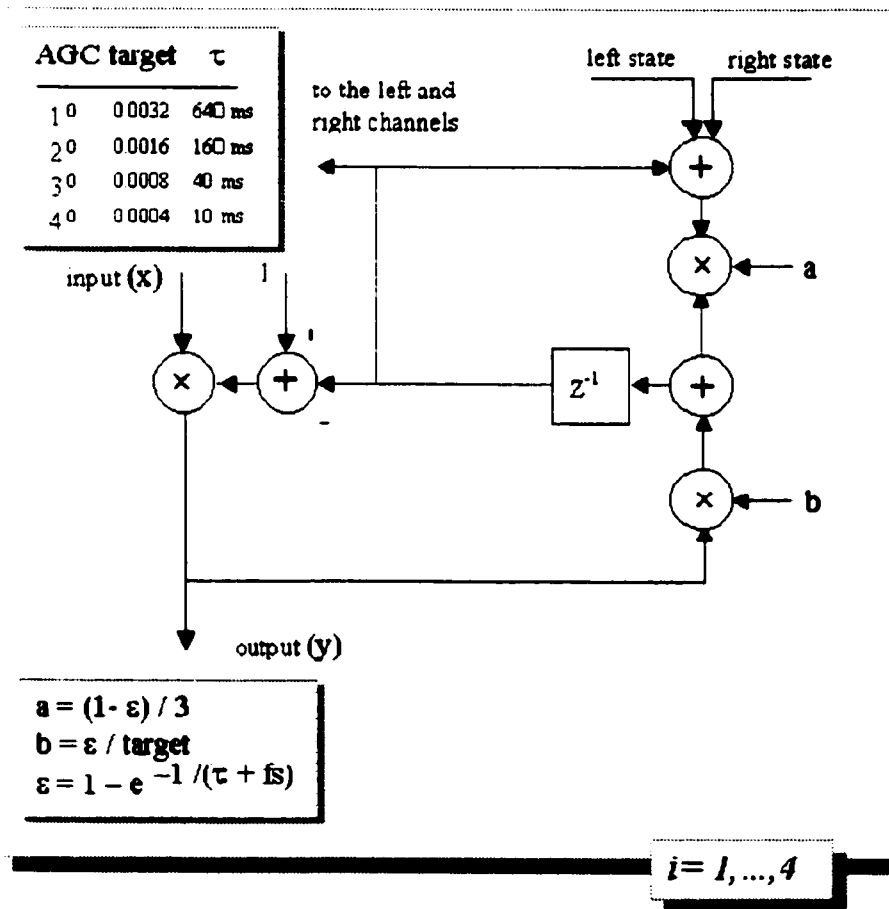


Fig 2.9 Block Diagram of AGC

Its state equation is given as:

$$\begin{aligned} \text{Gain}(n) &= \frac{y(n)}{x(n)} = 1 - w(n-1) \\ w(n+1) &= by(n+1) + a[w(n) + \text{left}(n) + \text{right}(n)] \end{aligned} \quad (2.30)$$

where :

$\text{left}(n)$ and $\text{right}(n)$ are the states of previous and next channel of the present channel in the cascaded model.

The main purpose of the AGC is to attenuate the incoming signal so that on average it remains below the target value t . In model the feedback with gain e / t is used to adjust the gain of the AGC with the output y . The steady state output of the AGC, y , is given as:

$$y \rightarrow \frac{t \times x}{t + x} \quad (2.31)$$

where: t is the target value of the AGC and x is the signal output from the second filter.

Figure 2.10 presents relationship of the output y and input x of the AGC. Note that the output of the AGC is dependent on both the input value and AGC target value.

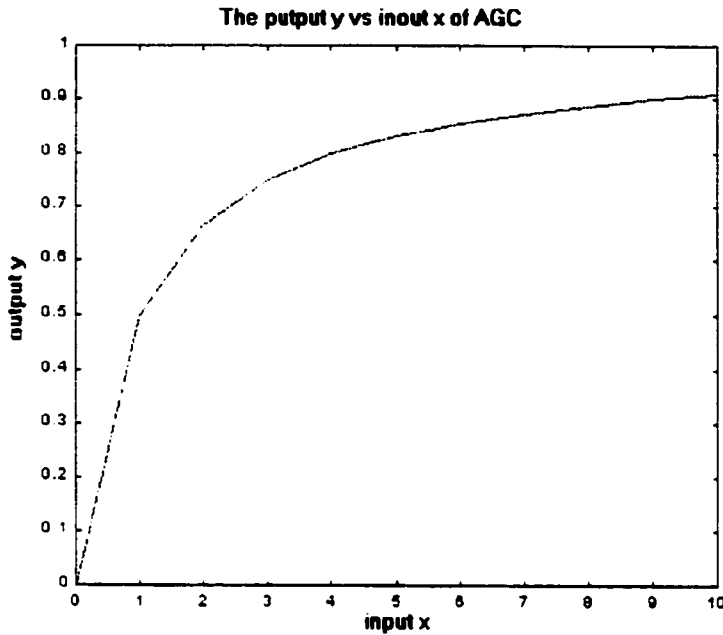


Fig 2.10 Output y vs. Input x of AGC when target = 1

The second purpose of AGC is to simulate the different adaptive times of the auditory system. In the model the loop with a feed back gain of $(1 - e) / 3$ represents a simple lowpass filter with a time constant related to parameter e . The target t and time constants τ for each stage in the model are given in table 2.2.

Table 2.2 Target t and Time Constant τ in the AGC Stages

Stage	Target t	Time Constant τ ms)
1 st AGC	0.0032	640
2 nd AGC	0.0016	160
3 rd AGC	0.0008	40
4 th AGC	0.0004	10

Figure 2.11 presents the outputs of AGC using different parameters. Note that the adaptive time is dependent on time constant τ while the steady state amplitude is dependent on both input and target t . The oscillatory behavior will be likely to happen when the time constant the AGC target get smaller.

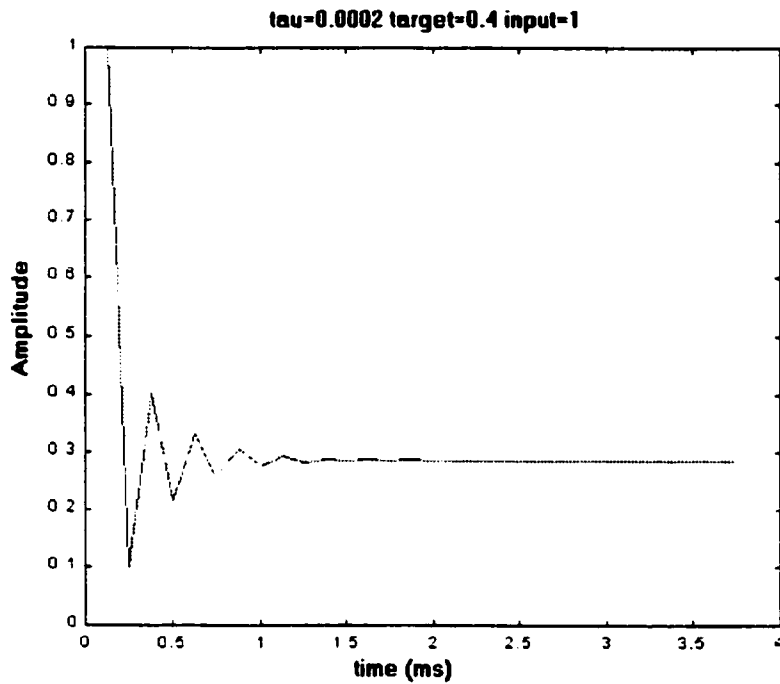


Fig. 2.11 Relationship of the AGC Output with Different τ and Target Values

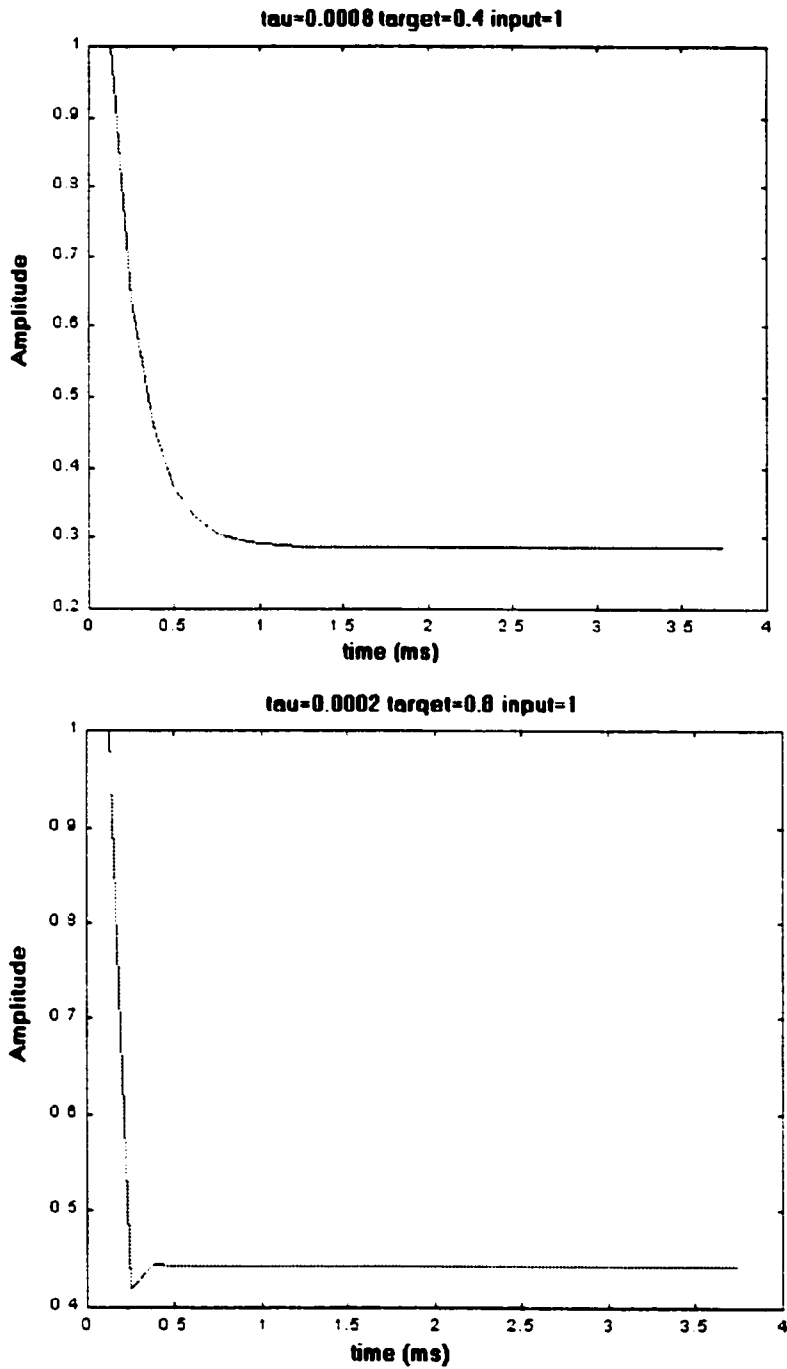


Fig . 2.11 Relationship of the AGC Output with Different τ and Target Values (con't)

The third purpose of AGC is to model the masking effects of the ear, i.e., the channel nonlinear stage output dependent not only on its own second order filter output but on its nearest channels. It is achieved by summation of its own state variable with state variables of its previous and next channels in gain $(1-e) / 3$ loop. Since all channels are coupled one channel can affect all channels in the filter bank although the effect will decay exponentially with distance.

9. Summary

This model represents the mechanical filtering of the mammalian auditory system; the model input is the sound signal at the outer ear, and output is the basilar membrane movement modeled as the filterbank outputs. By comparing with the experiment data the model provides the more accurately simulation results than traditional models.

Chapter 3

Ensemble Interval Histogram Model

Section 1 Introduction

Section 2 Cochlear Filterbank

Section 3 Level Crossing Detection

Section 4 Interval Histogram

Section 5 Summation of the EIH from All Channels

Section 6 The Choice of Bin Allocation over Frequency Range

Section 7 The Choice of the Window Length

1. Introduction

The ensemble interval histogram (EIH) model is schematically illustrated in figure 3.1. It includes 1. cochlear filterbank, 2. level crossing detection, 3. interval histogram calculation, and 4. summation of the histograms.

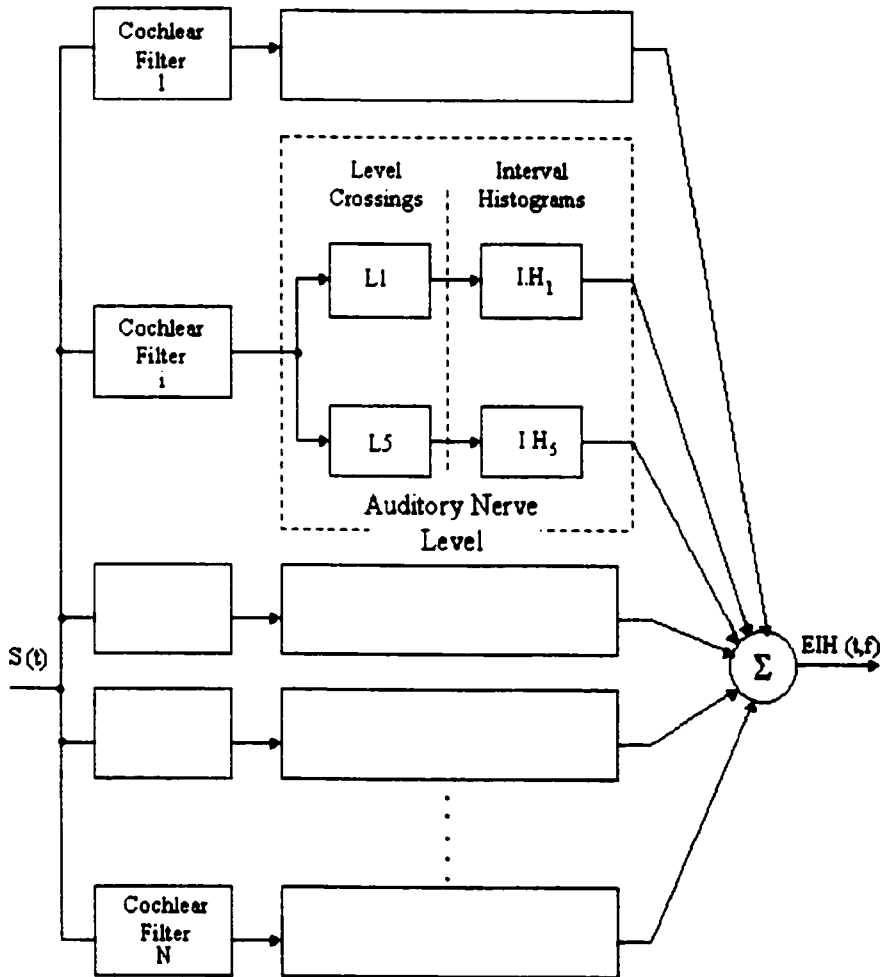


Fig 3.1 The Block Diagram of EIH

2. Cochlear Filterbank

In the EIH model, the mechanical motion of the basilar membrane is sampled using filterbank that is introduced in previous section. The number of channels is based on the speech signal sample rate [Ghitza94].

3. Level Crossing Detection

The next stage of processing in the figure 3.1 is an array of level crossing detectors that models the motion-to –neural activity transduction of the hair cell mechanisms. The detection levels of each detector are equally distributed on a log scale, thereby simulating the variability of fiber diameters and their synaptic connections.

4. Interval Histogram

Conceptually, the EIH is a measure of the spatial extent of coherent neural activity across the simulated auditory nerve. Mathematically, it is the short-term probability density function of the reciprocal of the intervals between successive firings, measured over the entire simulated auditory nerve in a CF dependent time frequency zone.

In the model, an estimate of the interval probability density function of a given level can be obtained by computing a histogram of the intervals from the point process data produced by the level-crossing detector. Only intervals between successive upward-going level crossings are considered. Since we prefer an auditory representation in the frequency domain, the histogram of the inverse intervals is computed. This is computed by distributing the reciprocal of the intervals

in a histogram. The EIH is measured by summing the corresponding histogram bins across all levels and all channels into one histogram.

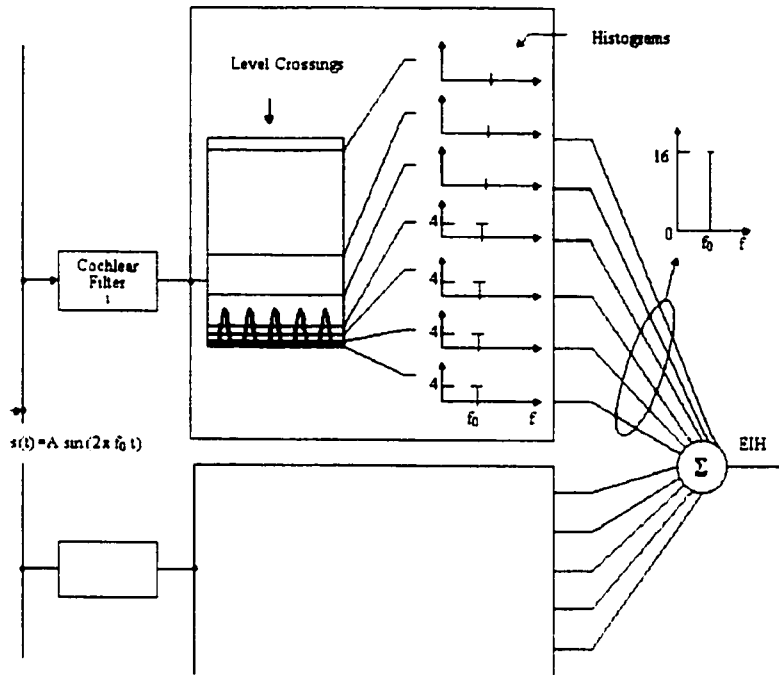


Fig 3.2 The Illustration of EIH Calculation of the i th Channel

The output of the level-crossing detectors represents the discharge activity of the auditory nerve fibers and it also preserves the signal overall energy information. To illustrate this point, consider the case in which the input signal is a pure sinusoid, i.e. $s(t) = A \sin(2\pi f_0 t)$, and the characteristic frequency of a selected channel is f_0 , as shown in figure 3.2. For a given intensity A , the output of the cochlear filter will activate only some low level-crossing detectors. For a given detector, the time interval between two successive positive-going level crossings is $1/f_0$. Since we prefer an auditory representation in the frequency domain, the histogram of the inverse intervals is computed, therefore, this interval contributes a count to the f_0 bin.

5. Summation of the EIH from All Channels

Figure 3.3 shows an input signal $s(t) = A \sin(2\pi f_0 t)$ driving five adjacent cochlear filters. Since the filters are linear we assume the filter amplitude response is $|H_i(f)|$ and the phase response is $\phi_i(f)$, $i = 1, 2, \dots, 5$. Due to the shape of the filters, more than one cochlear filter will contribute to the f_0 bin.

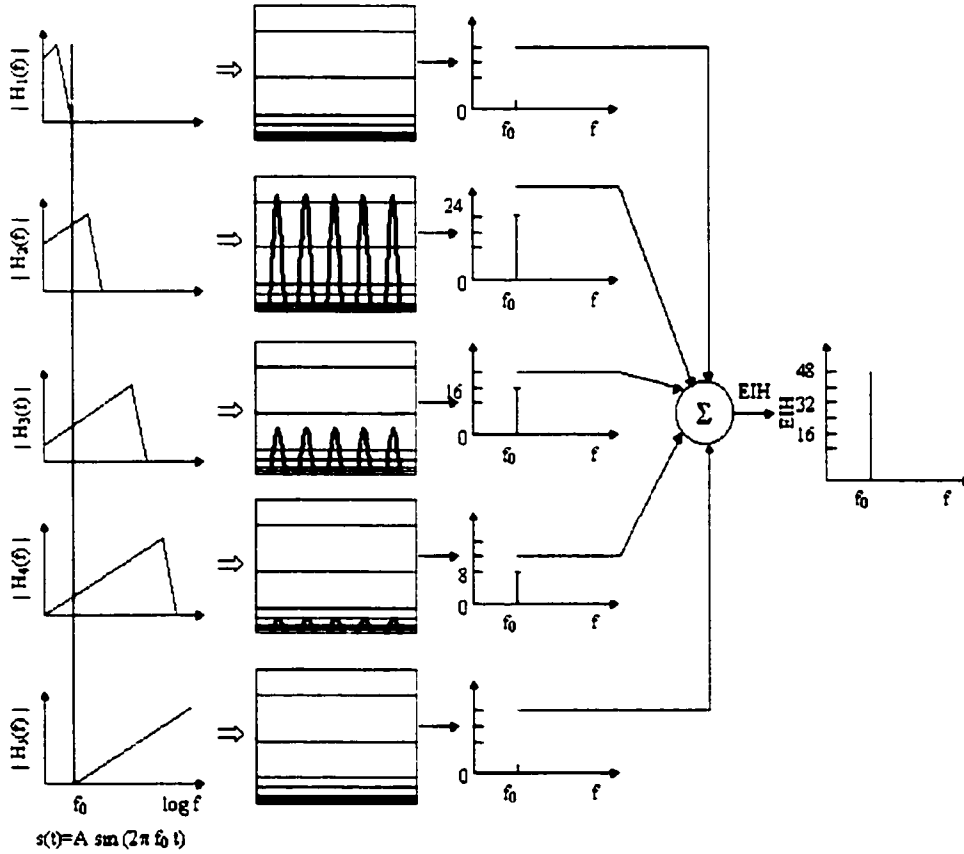


Fig 3.3 The Illustration of EIH Summation from 5 Successive Channels

6. The Choice of Bin Allocation over Frequency Range

Motivated by the tonotopic organization along the auditory pathway the bins are distributed according the ERB-rate scale. Let $|H(f)|$ be a frequency response of a filter, and let $|H(f_0)|$ be the maximum gain of the filter, at frequency f_0 . The equivalent rectangular bandwidth (ERB, in Hz) of $|H(f)|$ is defined as follows:

$$\text{ERB} = \frac{\int |H(f)|^2 df}{|H(f_0)|^2} \quad (3.1)$$

Using the ERB of the auditory filter as a unit of measurement, Moore and Glasberg [Moore and Glasberg 83] suggested the ERB-rate scale which relates number of ERB's to frequency. This scale is given as:

$$\text{ERB-rate} = 11.17 \ln \frac{F + 0.312}{F + 14.675} + 43.0 \quad (3.2)$$

where F is frequency in kHz. ERB-rate determines the number of successive ERB's that covers the frequency range $[0, F]$.

Using the ERB-rate scale, we quantized the frequency range [0, 8000] Hz into 32 bins, which is roughly the number of ERB's in this frequency range. Figure 3.4 shows the ERB-rate as a function of frequency.

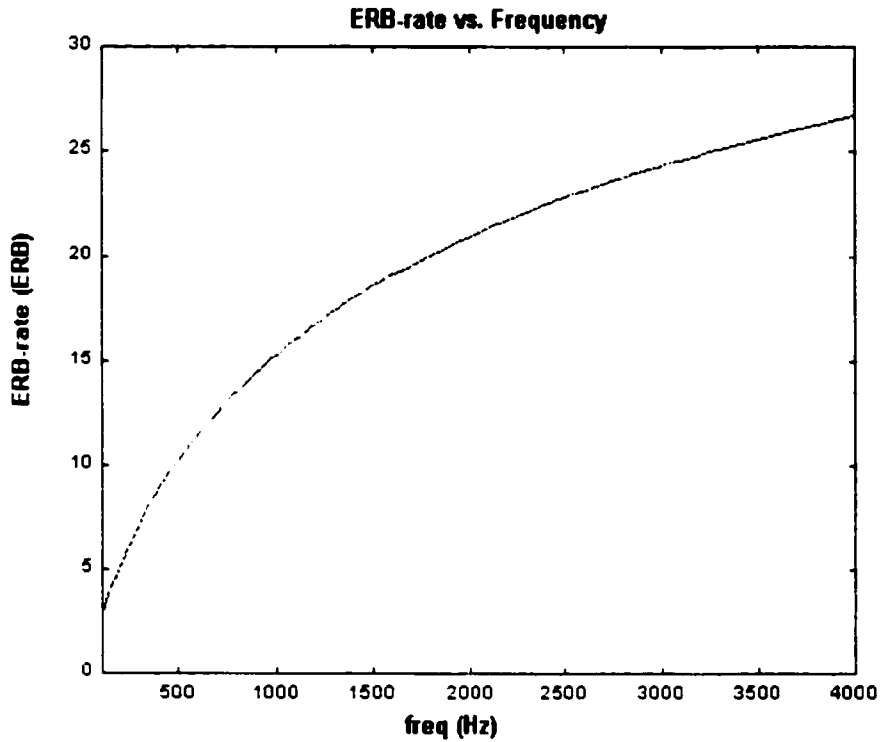


Fig 3.4 ERB-rate as a Function of Frequency.

7. The Choice of the Window Length

Motivated by the tonotopic organization along the auditory pathway again, the window length is set to be inversely proportional to center frequency. That is, at time t_0 , intervals produced by a level-crossing detector located at center frequency CF_0 are collected over a windows of length $10/CF_0$ that ends at time t_0 (see figure 3.5).

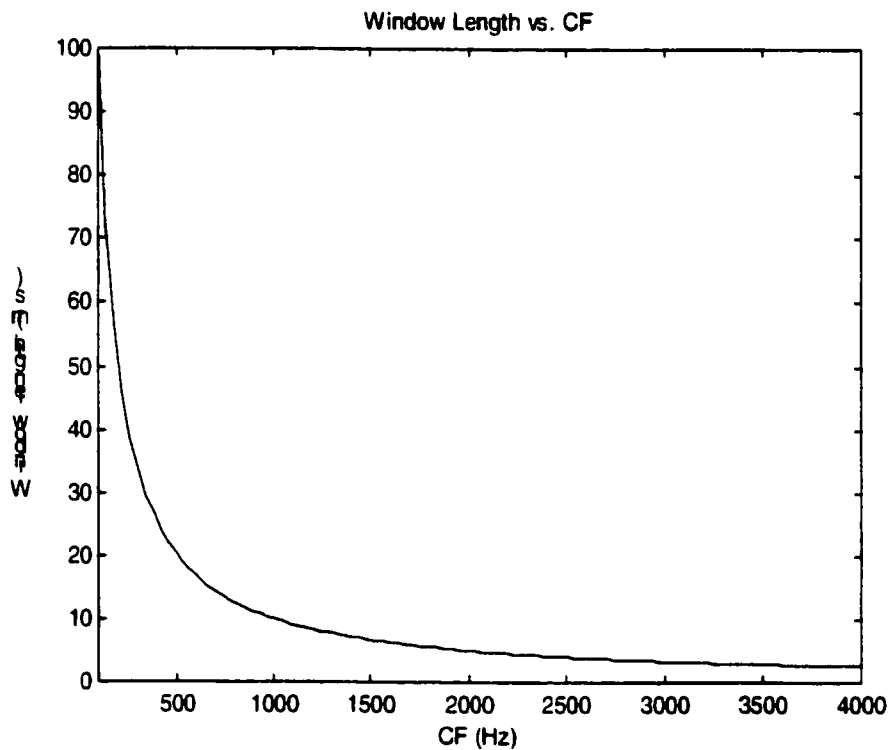


Fig 3.5 EIH Window Length as the Function of CF

C h a p t e r 4

Experiment Results

Section 1 Introduction

Section 2 Filterbank Simulation Results and Analysis

Section 3 EIH Simulation Results and Analysis

Section 4 Speech Recognition Experiment Results

1. Introduction

In this chapter some experiment results and analysis are presented. Firstly, the simulation results of the filter bank is given. Meanwhile, the comparison with experiment data is also presented. Secondly, some EIH simulations and some related discussion are described. Finally, we will describe some speech recognition experiments. It will demonstrate some speech recognition application of cochlear model and EIH model.

2. Filterbank Simulation Results and Analysis

Some frequency response results of the filterbank that is described in chapter 2 are plotted in figure 4.1. For comparison, some experiment data are also presented.

It shows that the model provides the good agreement with physiological data. The important features that match in the physiological data and simulation are:

- A very steep high-frequency slope
- A response zero about one octave below the peak
- A flat tail extending to the low-frequency cutoff of the middle-ear transfer function
- The amplitude ratio from tip to tail is approximately 60 dB at the highest frequencies and decreases as the center frequency decreases.

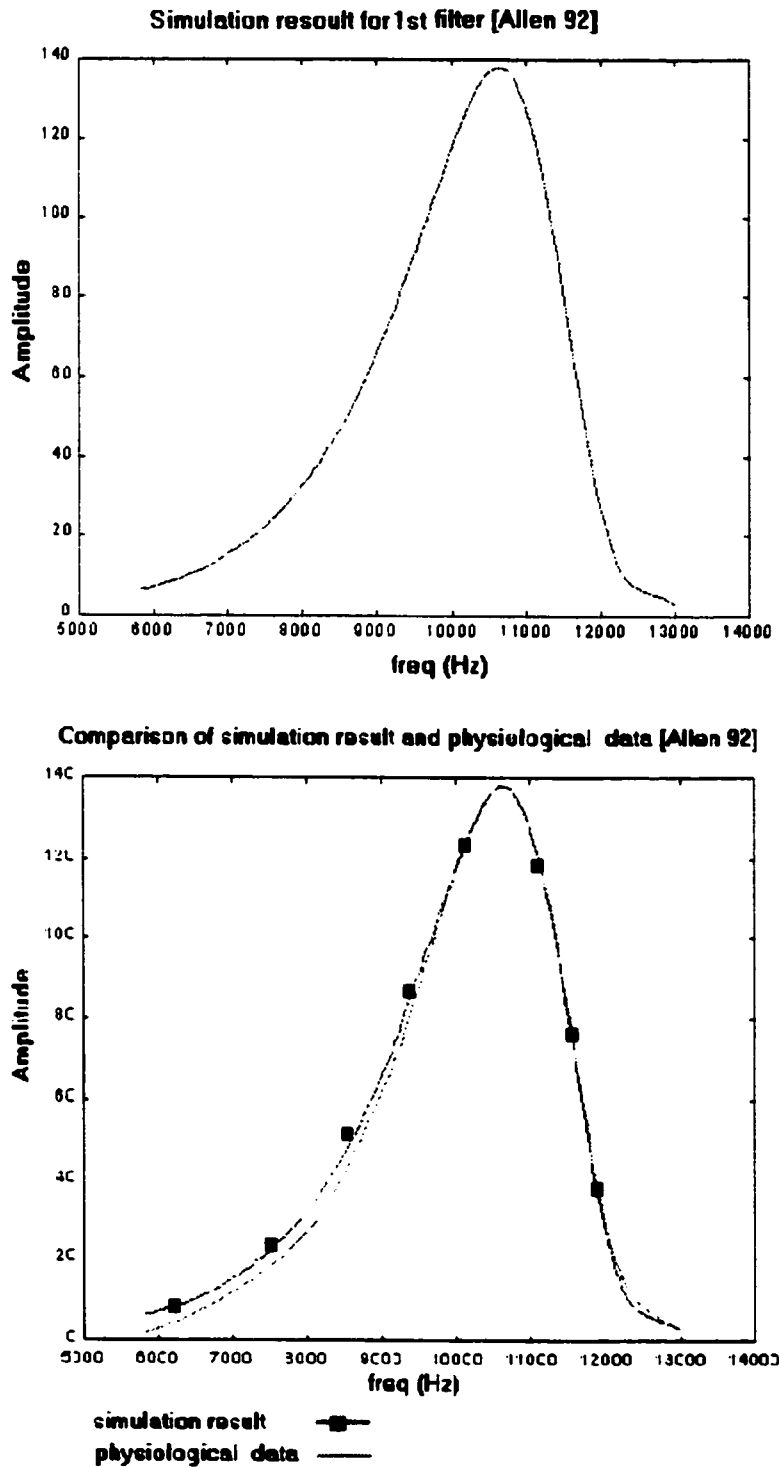


Fig 4.1(a) The Model Frequency Response at Tap 11600 (Hz) (Upper) and Comparison with Physiological Data

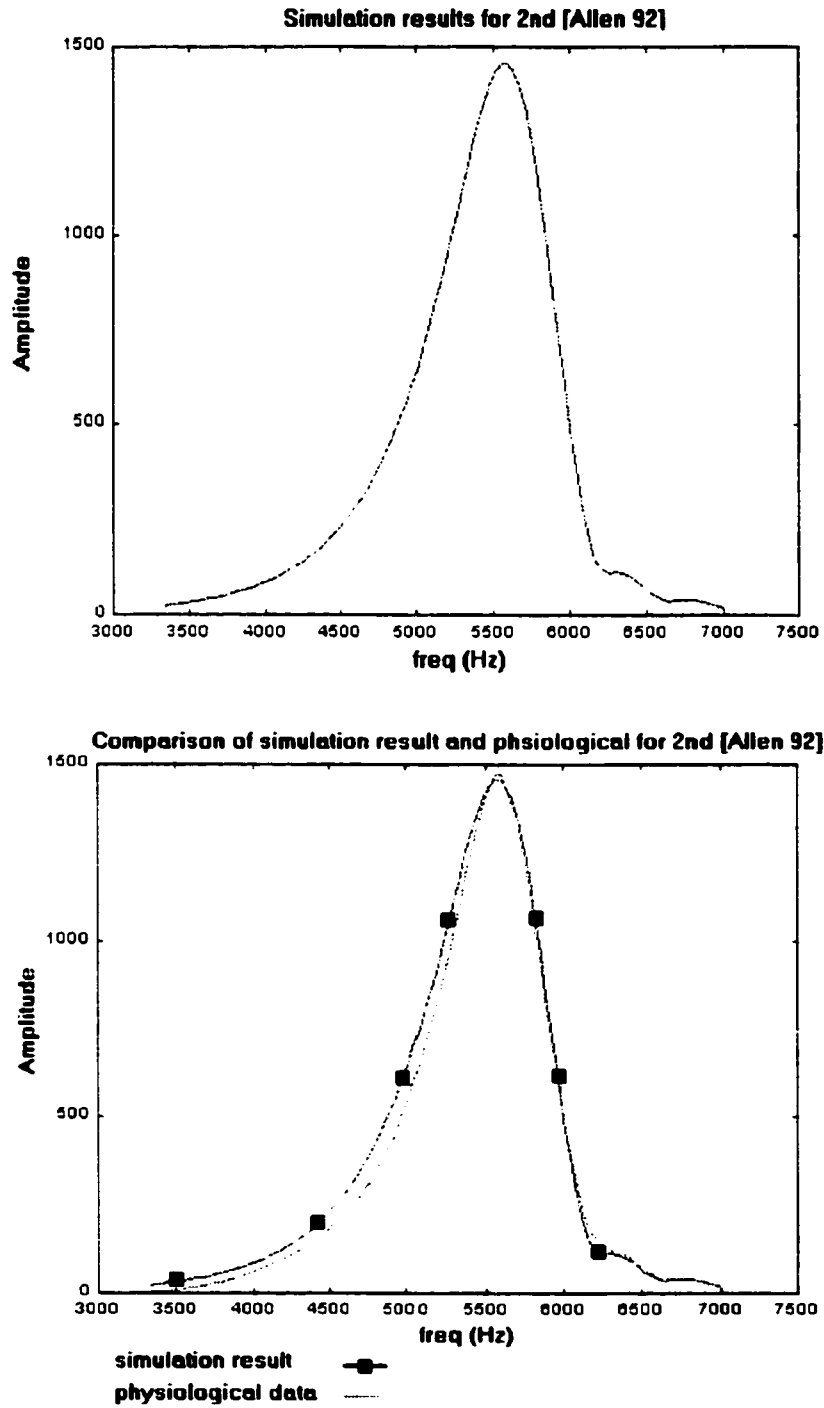
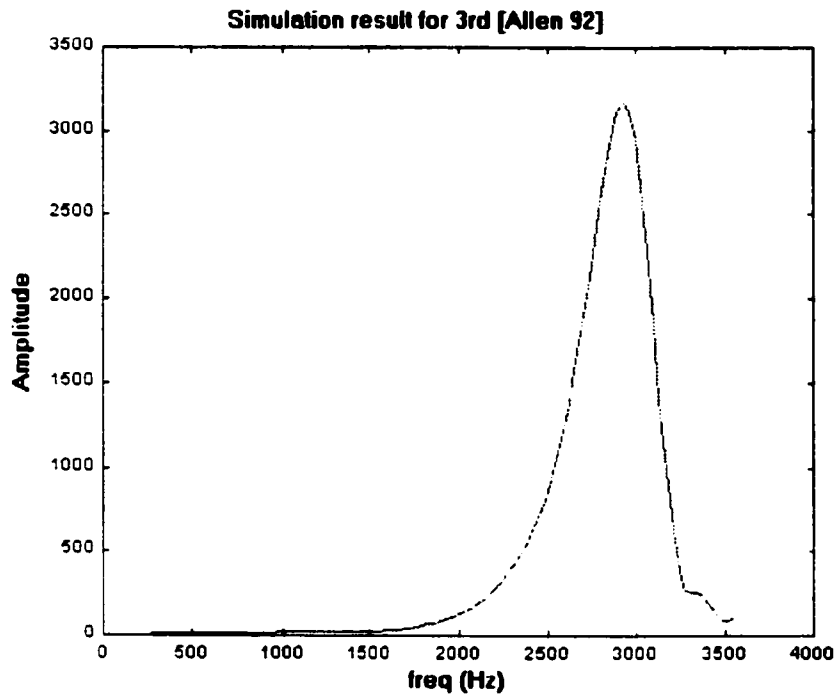


Fig 4.1(b) The Model Frequency Response (Blue) at Tap 5500 (Hz) (Upper) and Comparison with Physiological Data (Red)



Comparison of simulation result and physiological data 3rd filter [Allen 92]

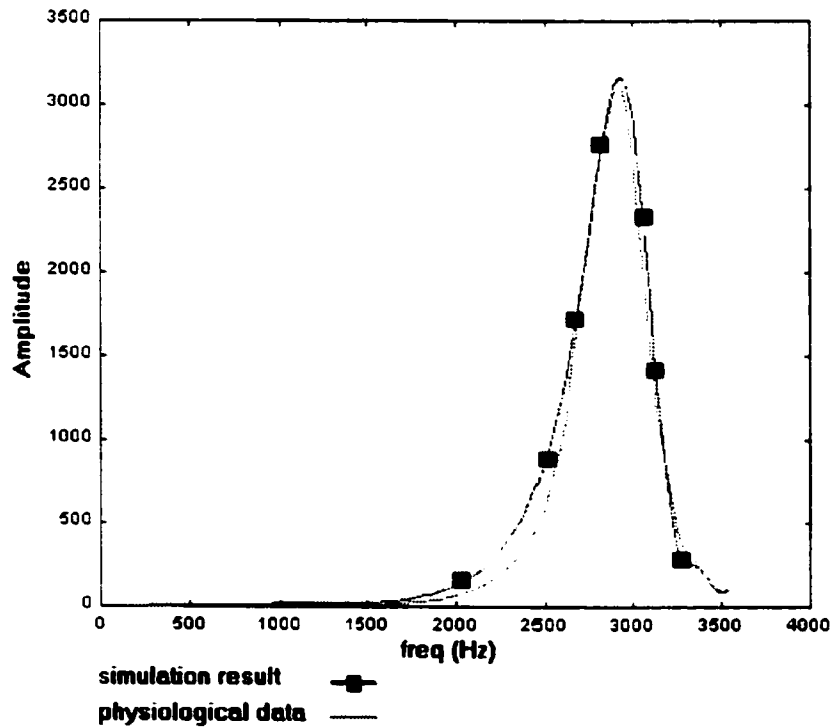


Fig 4.1(c) The Model Frequency Response (Blue) at Tap 3000 (Hz) (Upper) and Comparison with Physiological Data (Red)

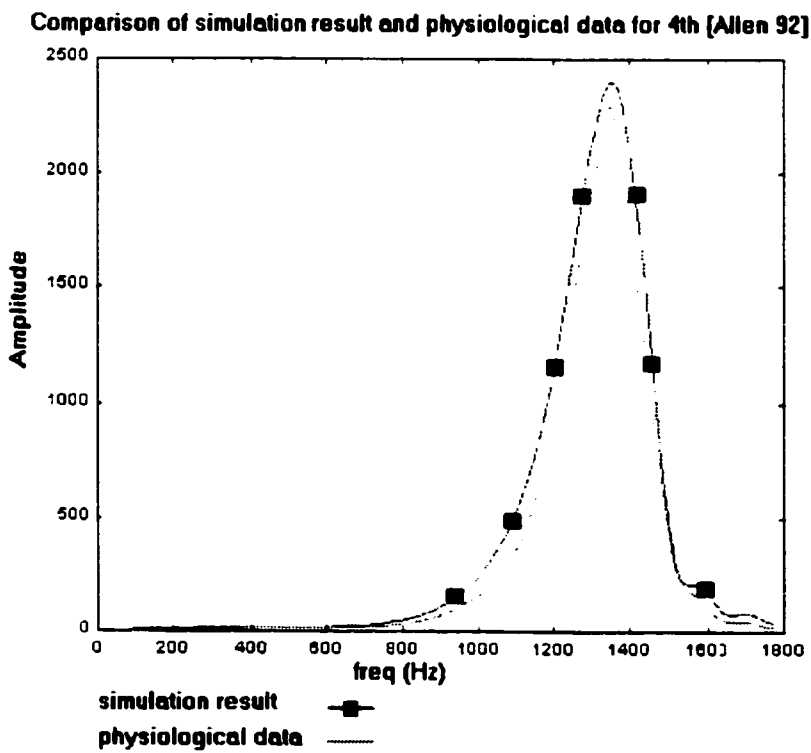
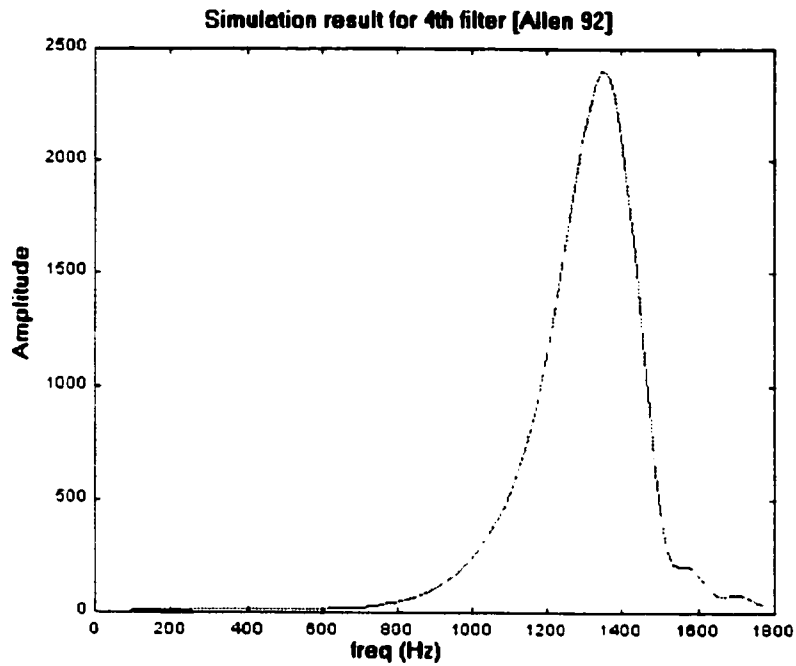
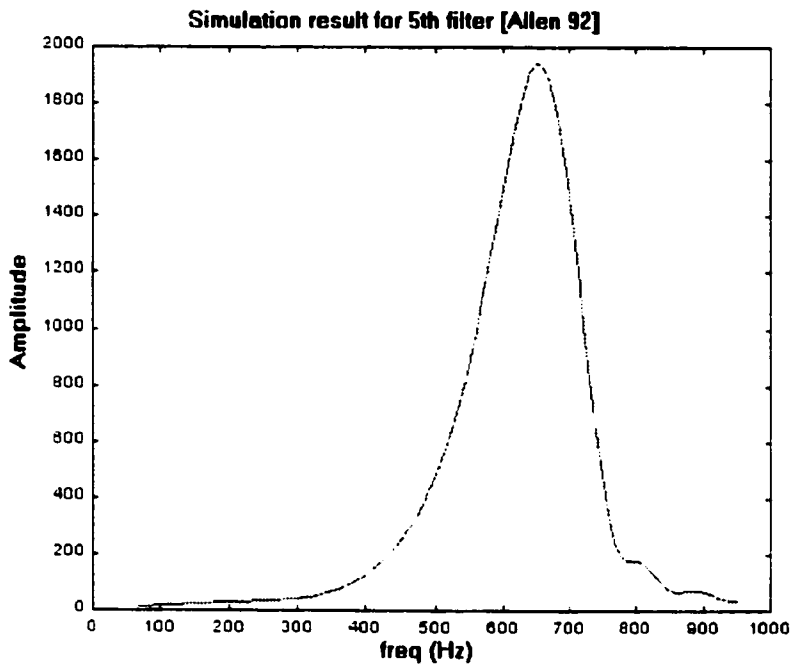


Fig 4.1(d) The Model Frequency Response (Blue) at Tap 1380 (Hz) (Upper) and Comparison with Physiological Data (Red)



Comparison of simulation result and physiological data for 5th filter [Allen 92]

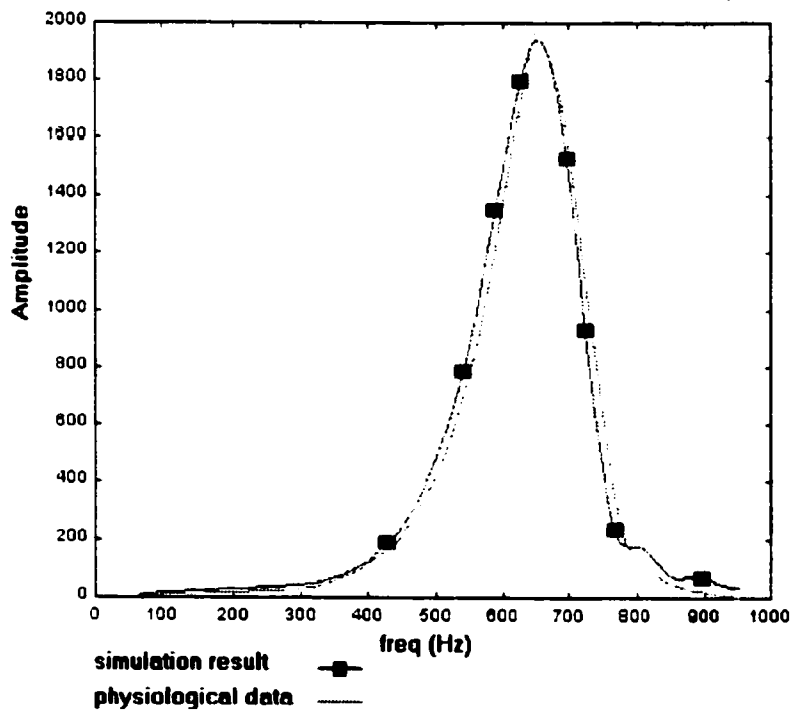
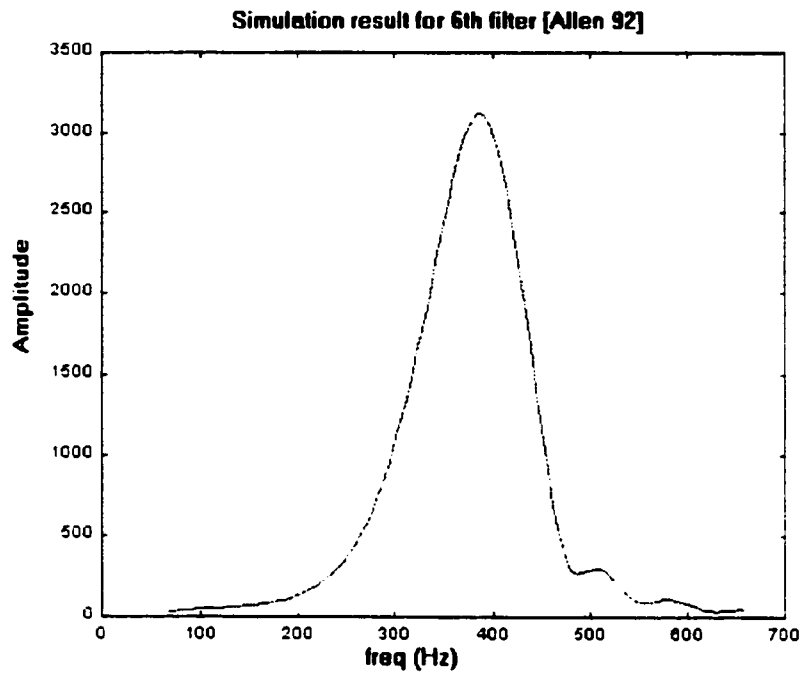


Fig 4.1(e) The Model Frequency Response (Blue) at Tap 650 (Hz) (Upper) and Comparison with Physiological Data (Red)



Comparison of simulation result and physiological data filter for 6th filter [Allen 92]

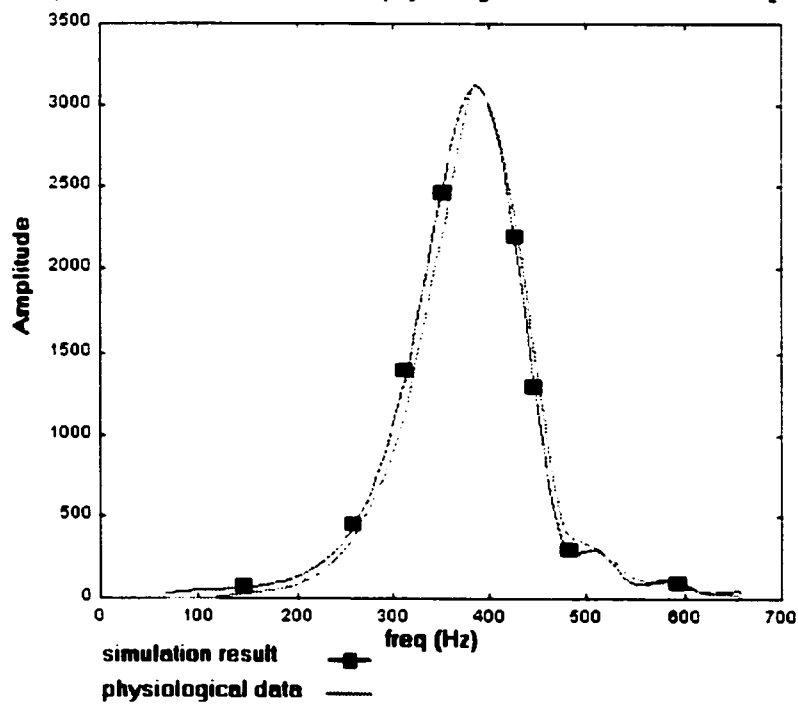


Fig 4.1(f) The Model Frequency Response (Blue) at Tap 400 (Hz) (Upper) and Comparison with Physiological Data (Red)

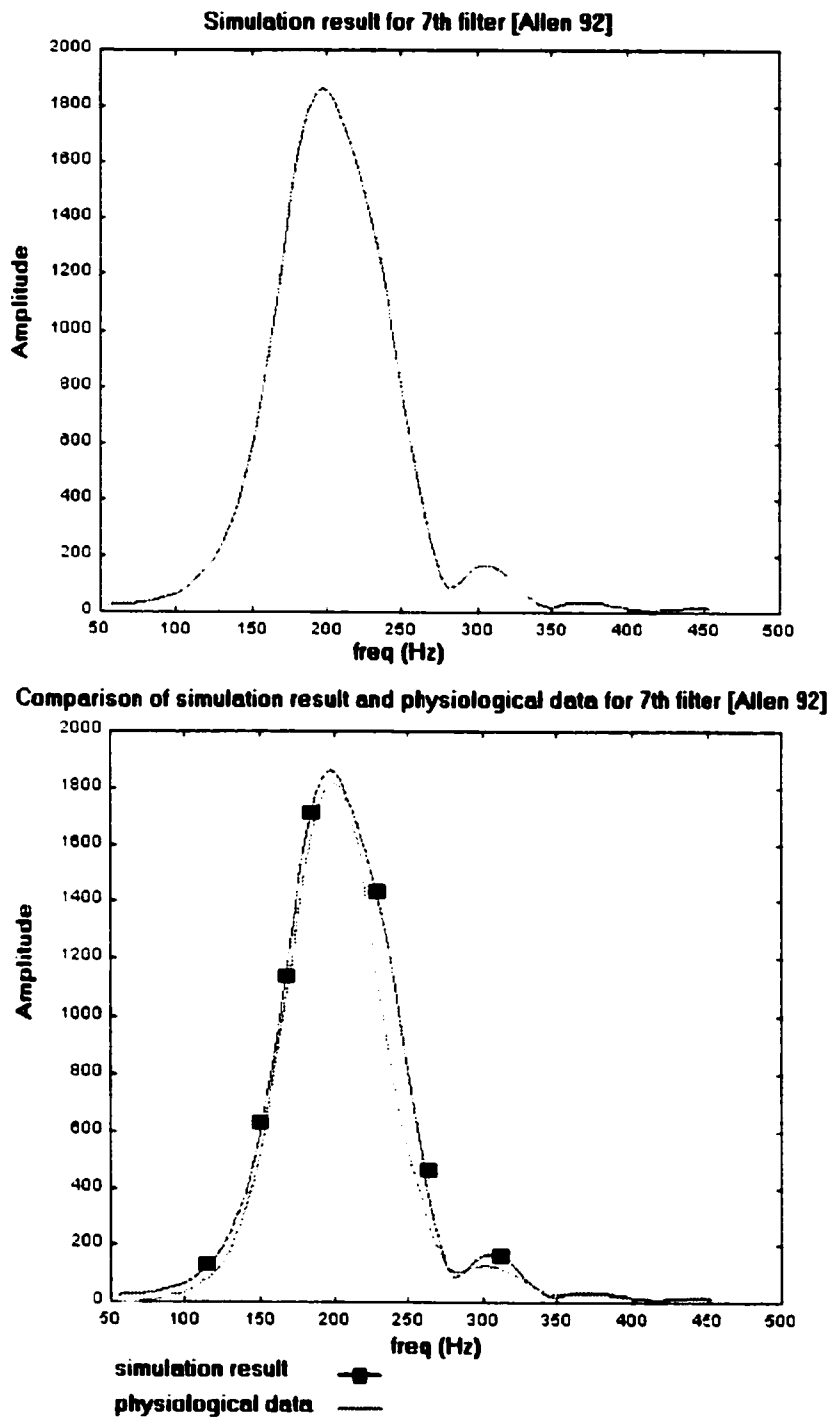


Fig 4.1(g) The Model Frequency Response (Blue) at Tap 200 (Hz) (Upper) and Comparison with Physiological Data (Red)

3. EIH Simulation Results and Analysis

This simulation used the cochlear model described in chapter 2 and the sample rate was 8kHz. The input signal was a pulse-train with 2ms pulse-width and 20ms period. The five threshold values were distributed evenly in log space. An EIH observation was computed once every 10 ms. The interval statistics at time t_0 were collected from all threshold detectors, using all simulated firing records which existed in the windows that ended at time t_0 . The length of each window was $10/CF$. The interval statistics were collected into 28-bin vector, where the frequency range [0 ,4000] Hz according to the ERB-rate scale.

Figure 4.2(a) illustrates the simulation results of the lowest CF channel in the cochlear model with CF 79Hz and bandwidth 125Hz. Figure 4.2(b) shows the simulation results of the highest CF channel in the cochlear model with CF 3811Hz and bandwidth 492Hz. The figures are organized from up to down with the increase of EIH collecting time. The EIH is produced at a uniform rate, once every 10 ms.

In figure 4.2(a) because the CF of the filter is at 79 Hz, the EIH followed this filter resolves the frequency component near 79 Hz. In contrast in figure 4.2(b), because the CF of the filter is at 3811 Hz the EIH of this filter resolves the high frequency component. Since the window length is changed with the CF, in low CF filter the EIH collecting window length is much longer than the frame length, then the interval histograms hardly change from frame to frame. In contrast, in high CF filter the EIH collecting window length is similar to the frame length, so the histograms change rapidly from frame to frame, demonstrating high temporal resolution.

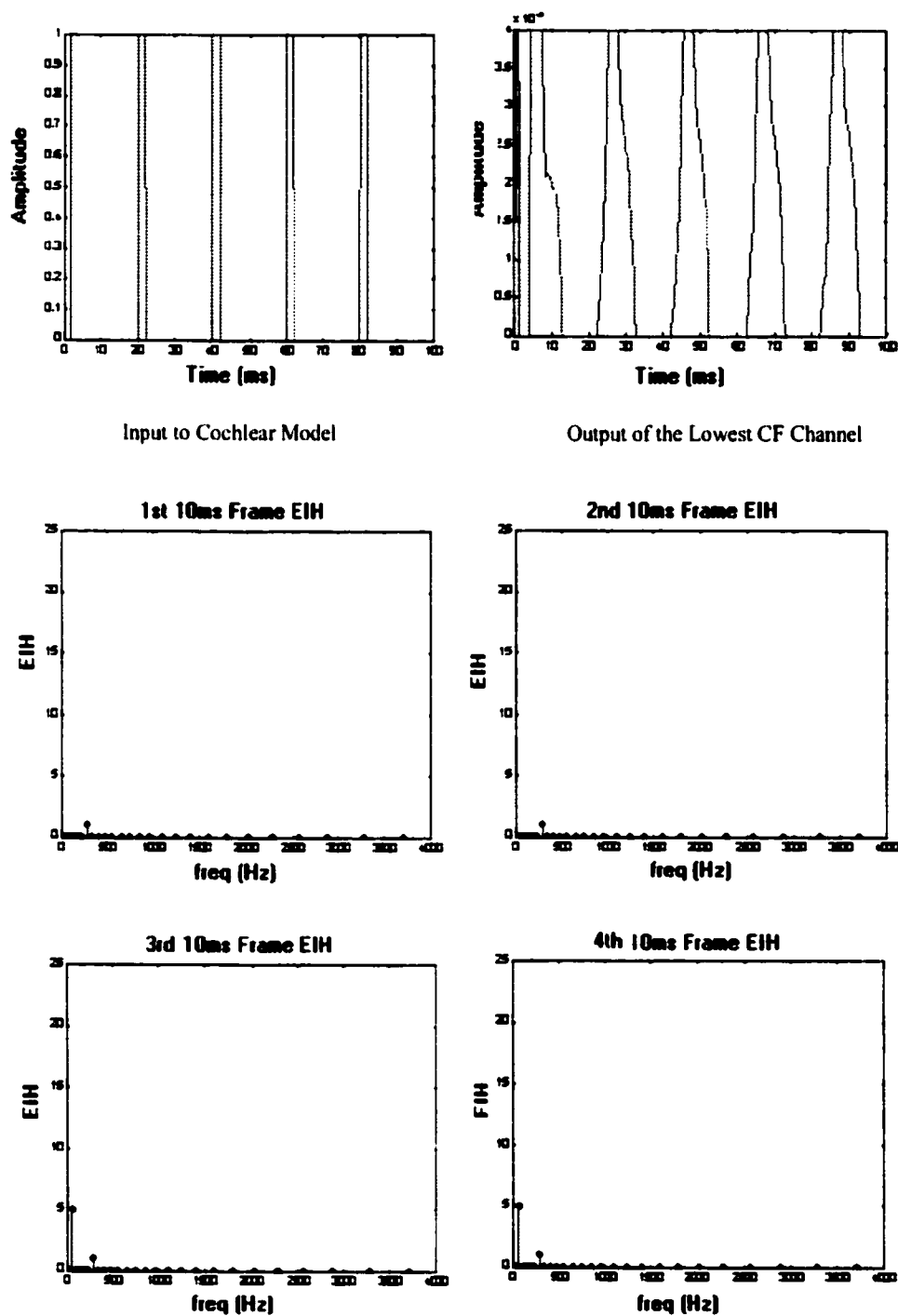


Fig 4.2(a) The EIH of the Lowest CF Channel in the Cochlear Model

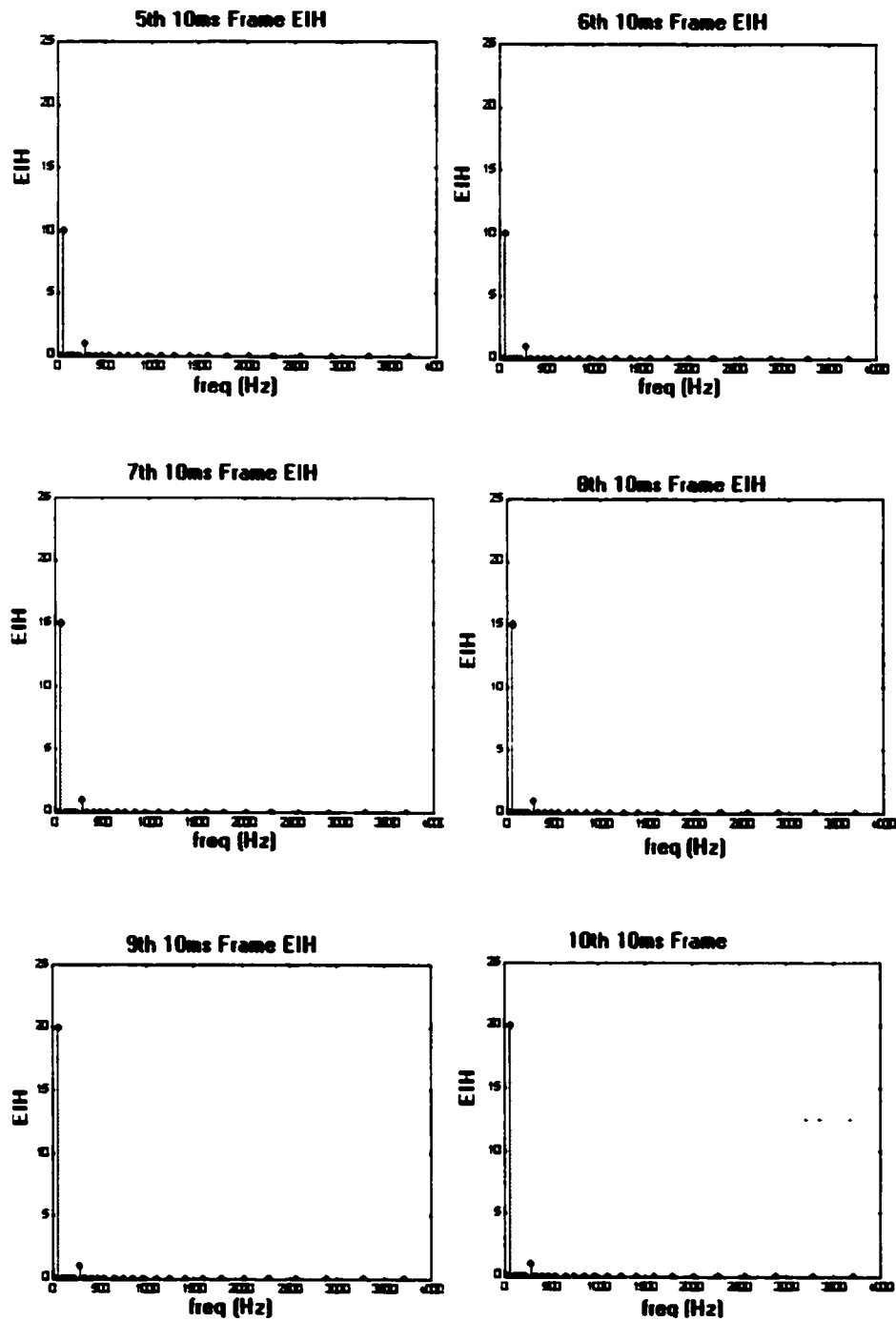
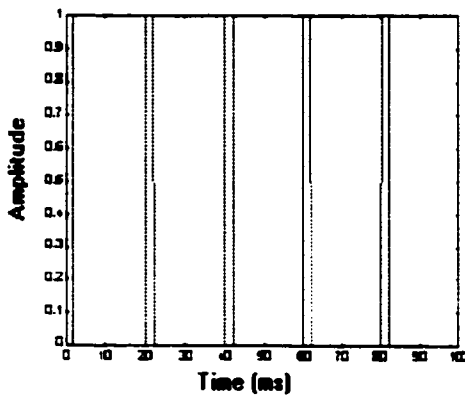
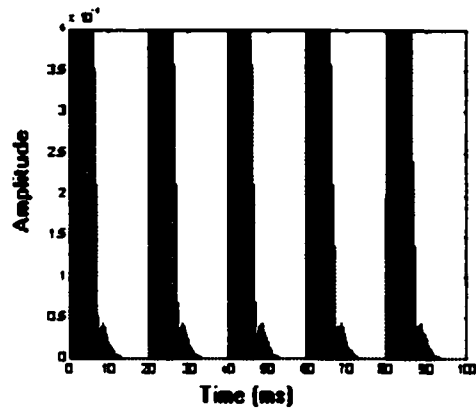


Fig 4.2(a) The EIH of the Lowest CF Channel in the Cochlear Model (cont'd)



Input to Cochlear Model



Output of the Highest CF Channel

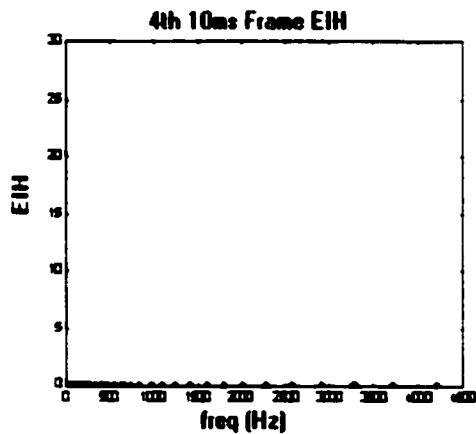
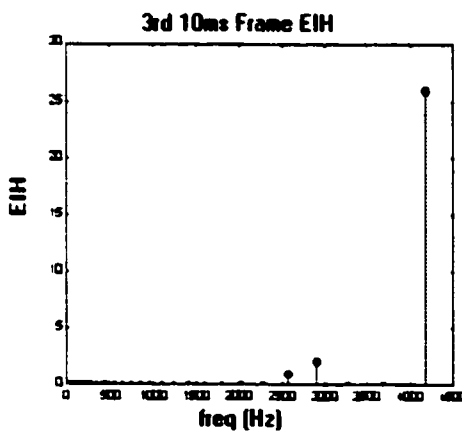
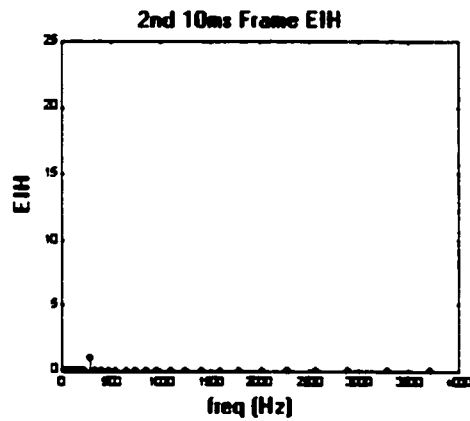
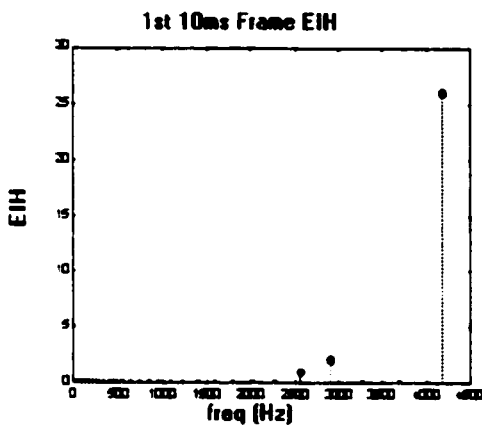


Fig 4.2(b) The EIH of the Highest CF Channel in the Cochlear Model

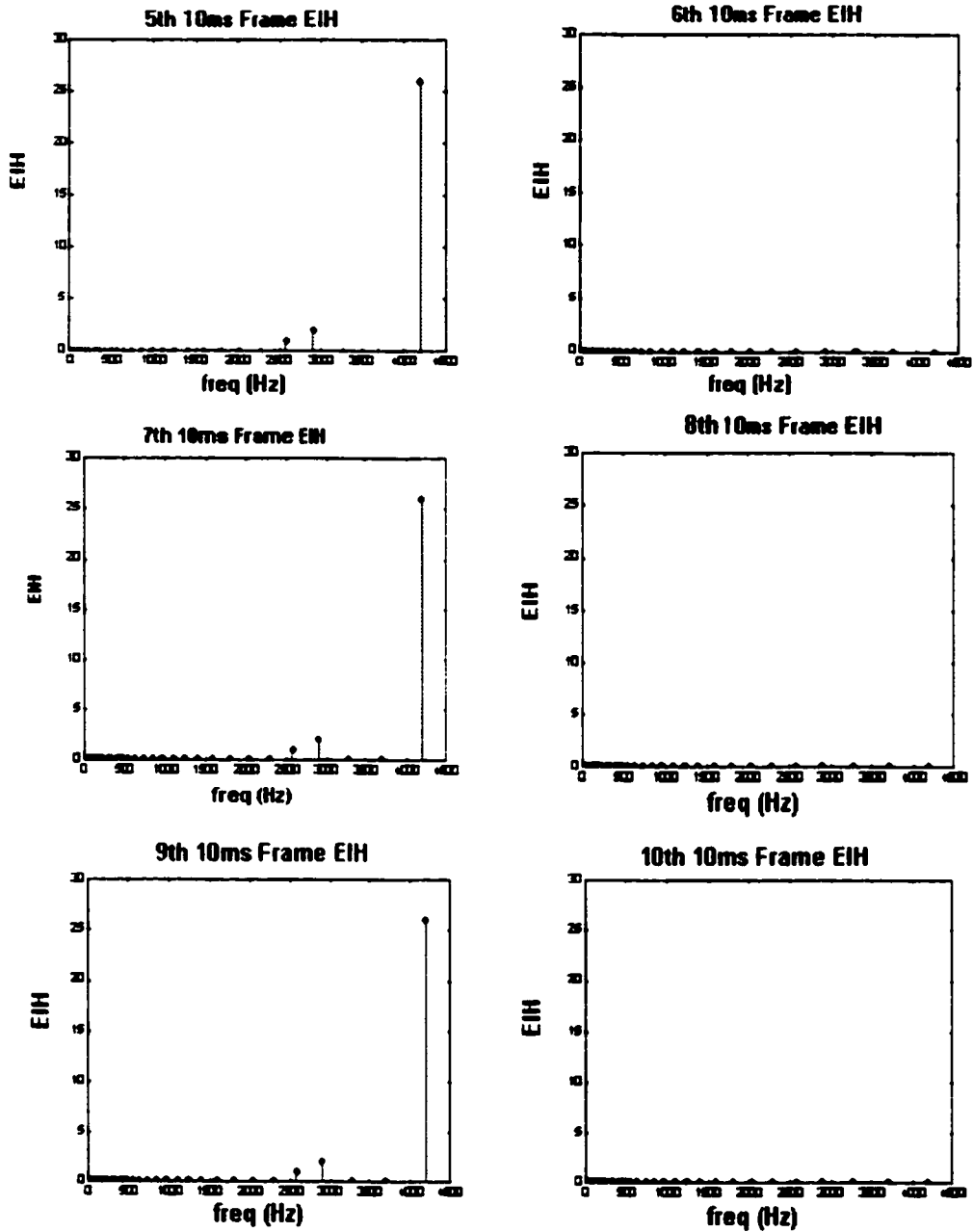


Fig 4.2(b) The EIH of the Highest CF Channel in the Cochlear Model (cont'd)

In figure 4.3 the linear bin allocation (128 bins over [0, 4000]) interval histograms are presented. In this case, since the filter bandwidth is related to the associated CF, in low CF interval histograms the bandwidth is narrow, so the EIH of this filter is identical to ERB allocation. But for high CF filter since the bandwidth is wider several bins are distributed over the interval histograms. However, in ERB case the bins at high frequencies are wide. Each ERB bin covers a wider frequency range that contains several linearly allocated bins. Therefore, the interval count at this ERB bin equals to the sum of intervals over all the linearly allocated bins at that frequency range. In other words, at time t_0 , ERB bins at high frequencies contain the overall number of intervals collected over the window, irrespective of the shape of the interval pdf. Therefore, we view the changes in time at the high frequency bins as a measure of the instantaneous rate.

From above results we could obtain the properties of the EIH, collected over several successive filters. For low CF filters they resolve low frequency components and form the related bins with the frequencies. Therefore, the magnitude of the EIH at these bins can be viewed as a measure of level-crossing detectors and the number of the successive filters that are synchronized (or phase-locked) to their underlying frequency components. For the high CF filters the changes with time of the corresponding EIH bin can be viewed as a measure of the coherent instantaneous-rate activities across the level-crossing detectors. These results follow the measurements of firing response of cats' auditory fibers ([Delgutte and Kiang 84a], [Sachs and Young 79]).

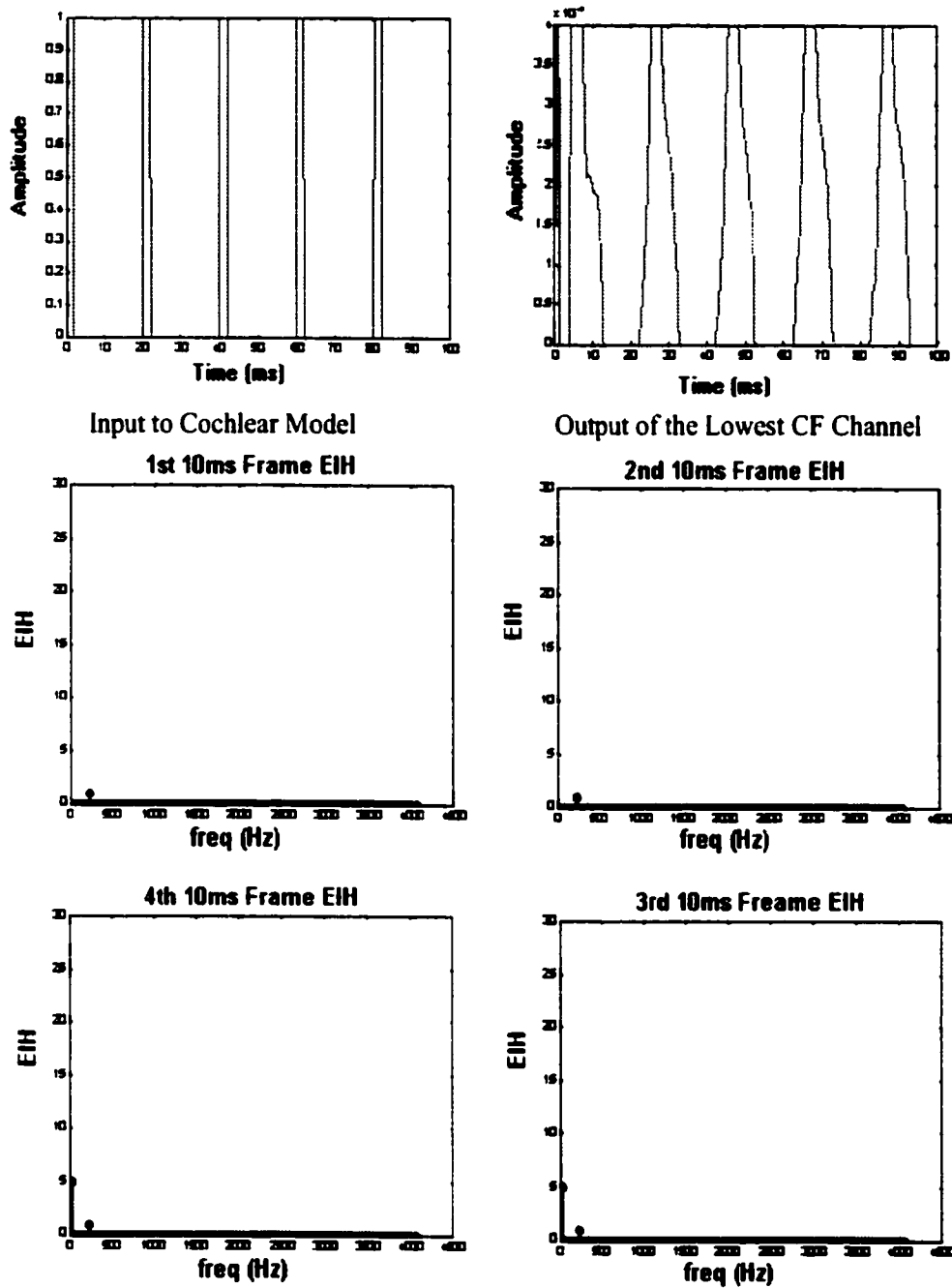


Fig 4.3(a) The Linear Bin Allocated EIH of the Lowest CF Channel in the Cochlear model

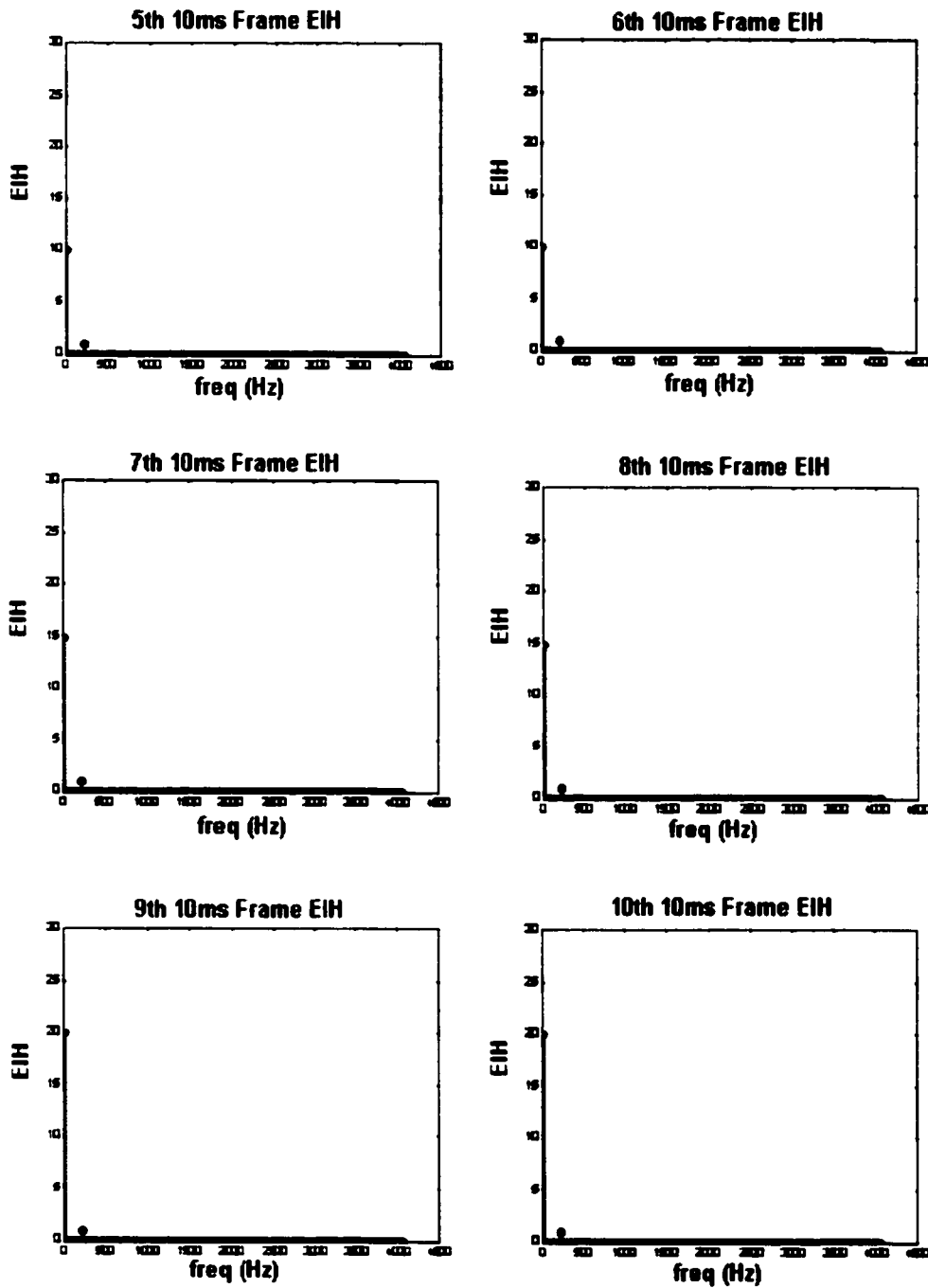
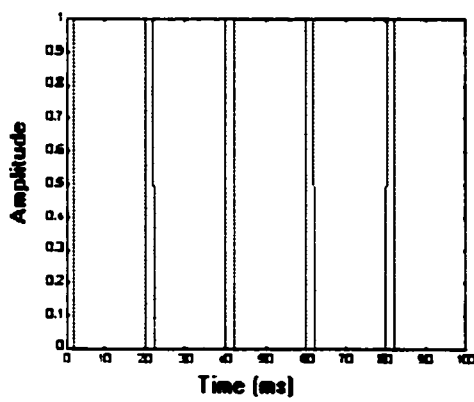
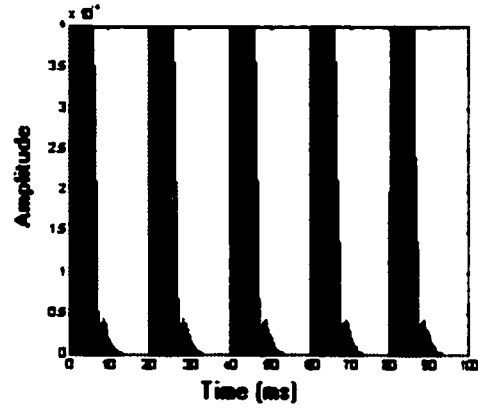


Fig 4.3(a) The Linear Bin Allocated EIH of the Lowest CF Channel in the Cochlear Model

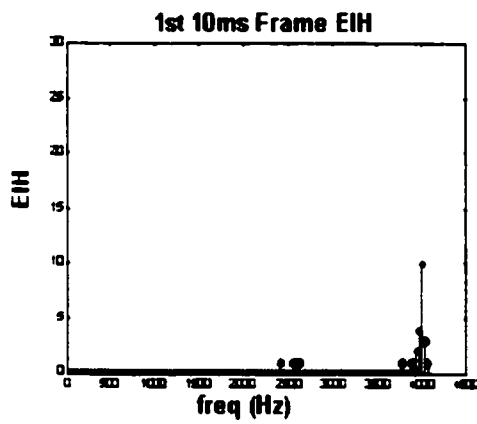
(cont'd)



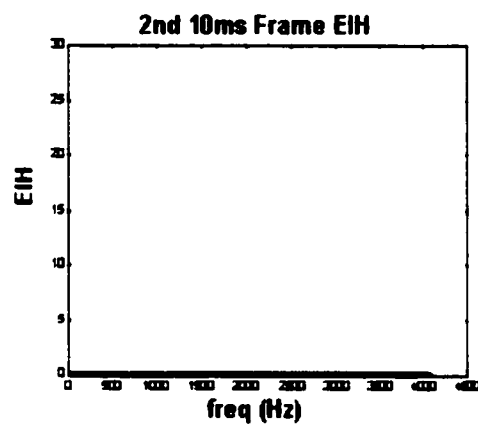
Input to Cochlear Model



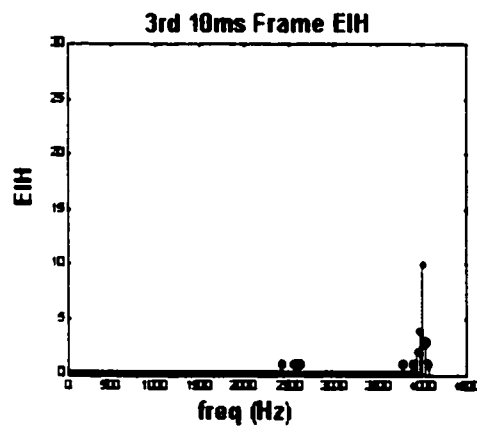
Output of the Highest CF Channel



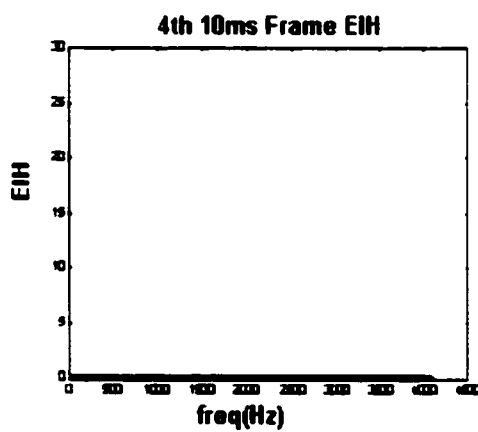
1st 10ms Frame EIH



2nd 10ms Frame EIH



3rd 10ms Frame EIH



4th 10ms Frame EIH

Fig 4.3(b) The Linear Bin Allocated EIH of the Highest CF Channel in the Cochlear Model

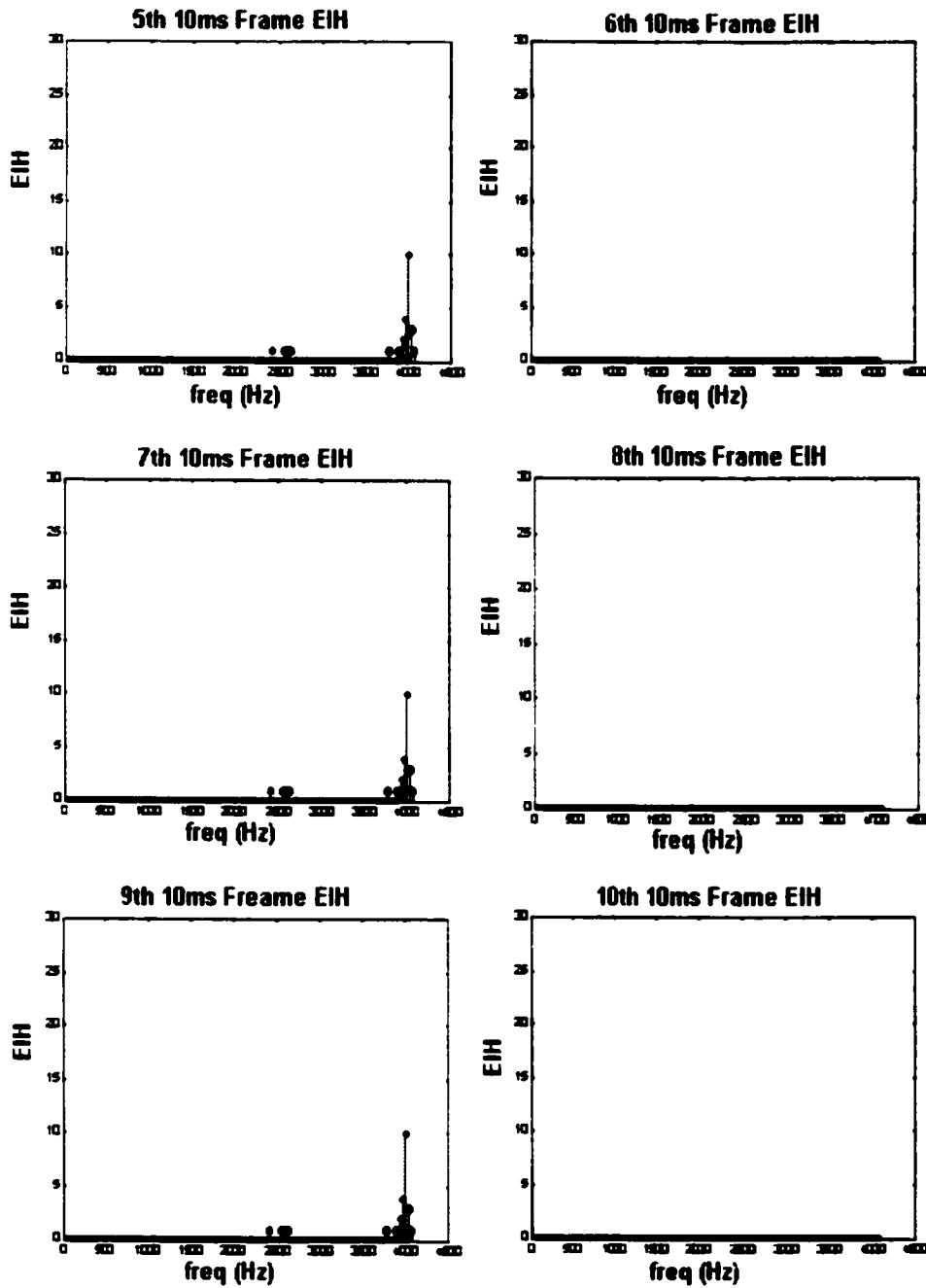


Fig 4.3(b) The Linear Bin Allocated EIH of the Highest CF Channel in the Cochlear Model

(cont'd)

4. Speech Recognition Experiment Results

4.1 Introduction

In this section we examine the extent to which the model we described in previous chapters is capable of performance to speech recognition. The entire recognition system includes cochlear model, EIH algorithms, and LVQ recognizer as shown in figure 4.4.

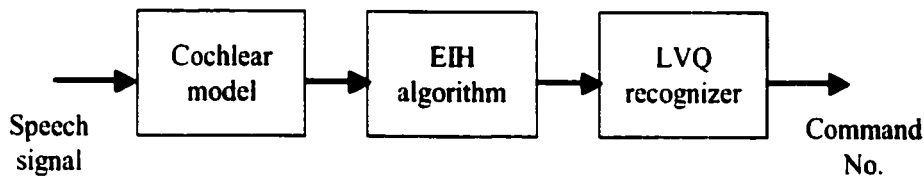


Fig. 4.4 Diagram of the Speech Recognition Experiment

The cochlear model and EIH algorithm are described in chapter 2 and chapter 3 respectively.

LVQ recognizer we used in the experiment is LVQ1 and LVQ3.

4.2 Signal Conditions

Three male speakers and three female speakers provided three repetitions of the 50 commands in the laboratory condition. The signals were sampled at 16 kHz rate and recorded as 16 bit per sample in windows wave files using sound recorder software.

4.3 Analysis Methods

We tested the EIH model by computing the EIH representation of each command. The auditory model for the EIH representation is described in chapter 2. An EIH observation was computed once every 10 ms from 32 thresholds per channel. The interval statistics at time t_0 are collected from all cochlear channel threshold detectors, using all simulated firing records which exit in the windows that end at times t_0 . The length of each window is $10/CF$, where CF is the characteristic frequency of the cochlear channel. The interval statistics are collected into a 128-bin vector, where the bins linearly divide the frequency range $[0, 4000]$ Hz. The sum over the 128 bins was normalized to 1 to eliminate the effect of “loudness”.

To demonstrate the effect of bin allocation and level crossing number to the recognition results, ERB bin allocation and 16 threshold level crossing are also used to calculate the EIH representation of the commands and applied to the recognizer to examine recognition results.

In order to examine the performance of the EIH model the LVQ recognizer is used to classify the 50 commands. We first used LVQ1 to classify the commands then used LVQ3 to improve the classification.

4.4 Experiment Results and Analysis

Figure 4.5 presents sound wave forms, cochleagrams, and EIH of 4 commands from 50 commands.

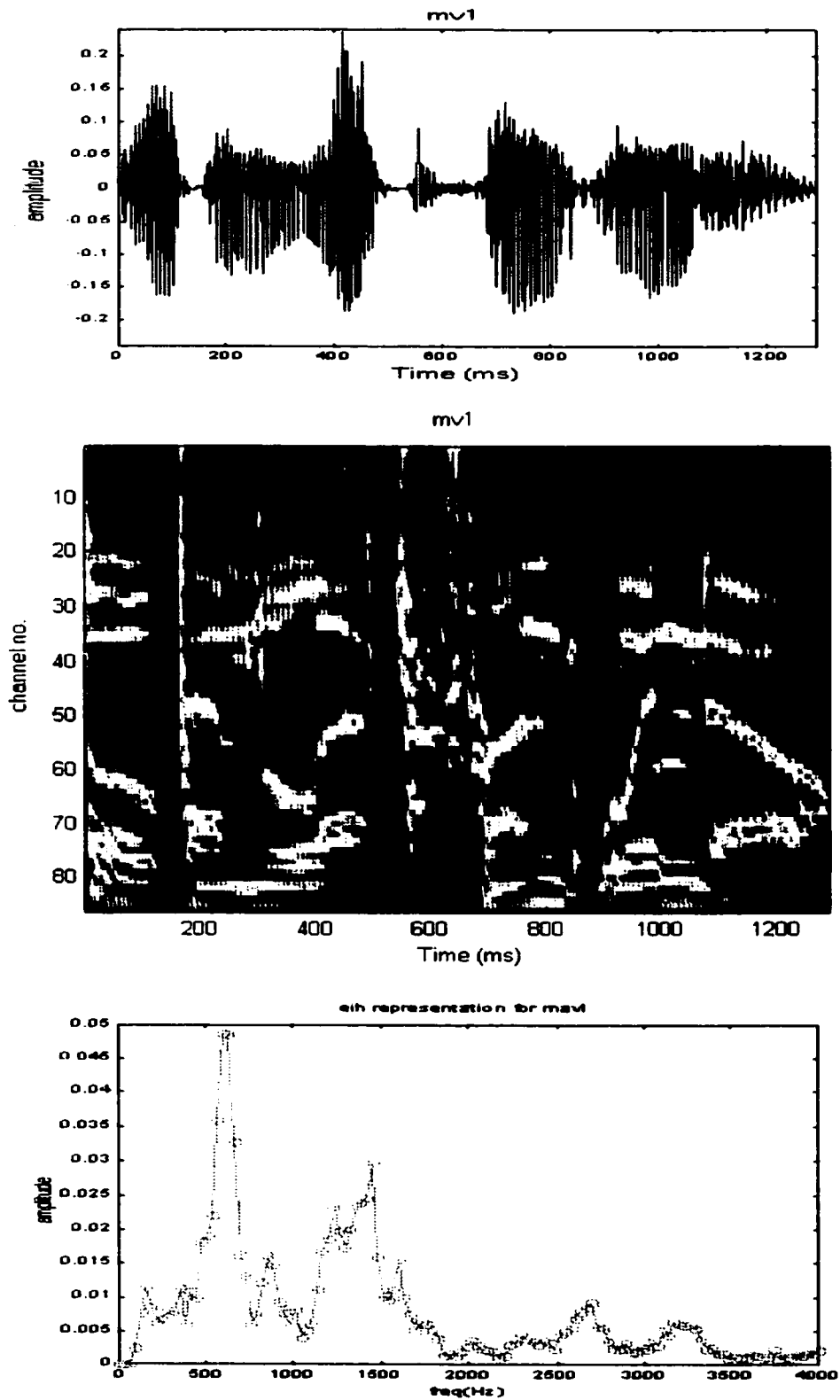


Fig 4.5(a) Sound Wave (upper), Cochleagram (middle), and EIH (below) of V1

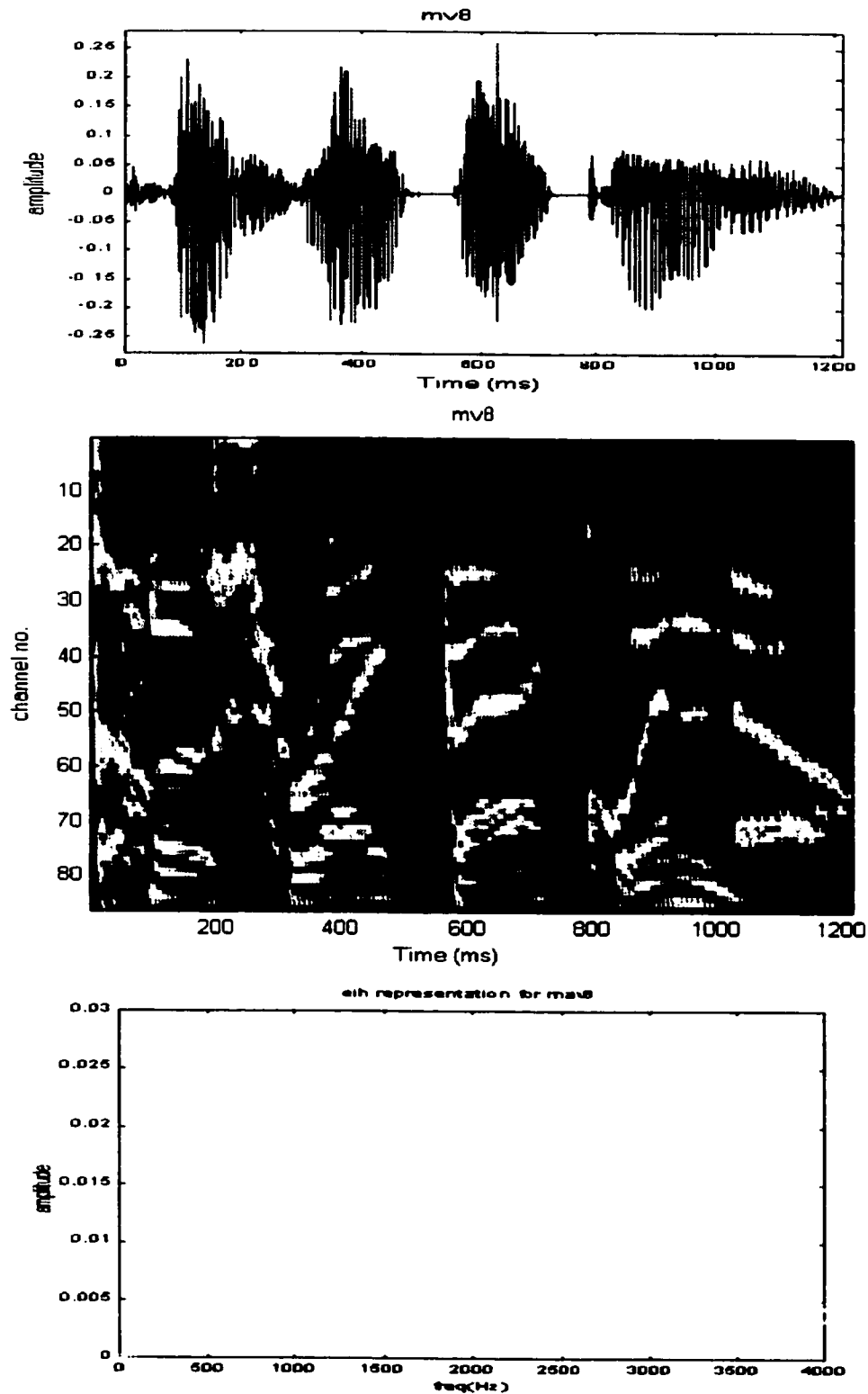


Fig 4.5(b) Sound Wave (upper), Cochleagram (middle), and EIH (below) of V8

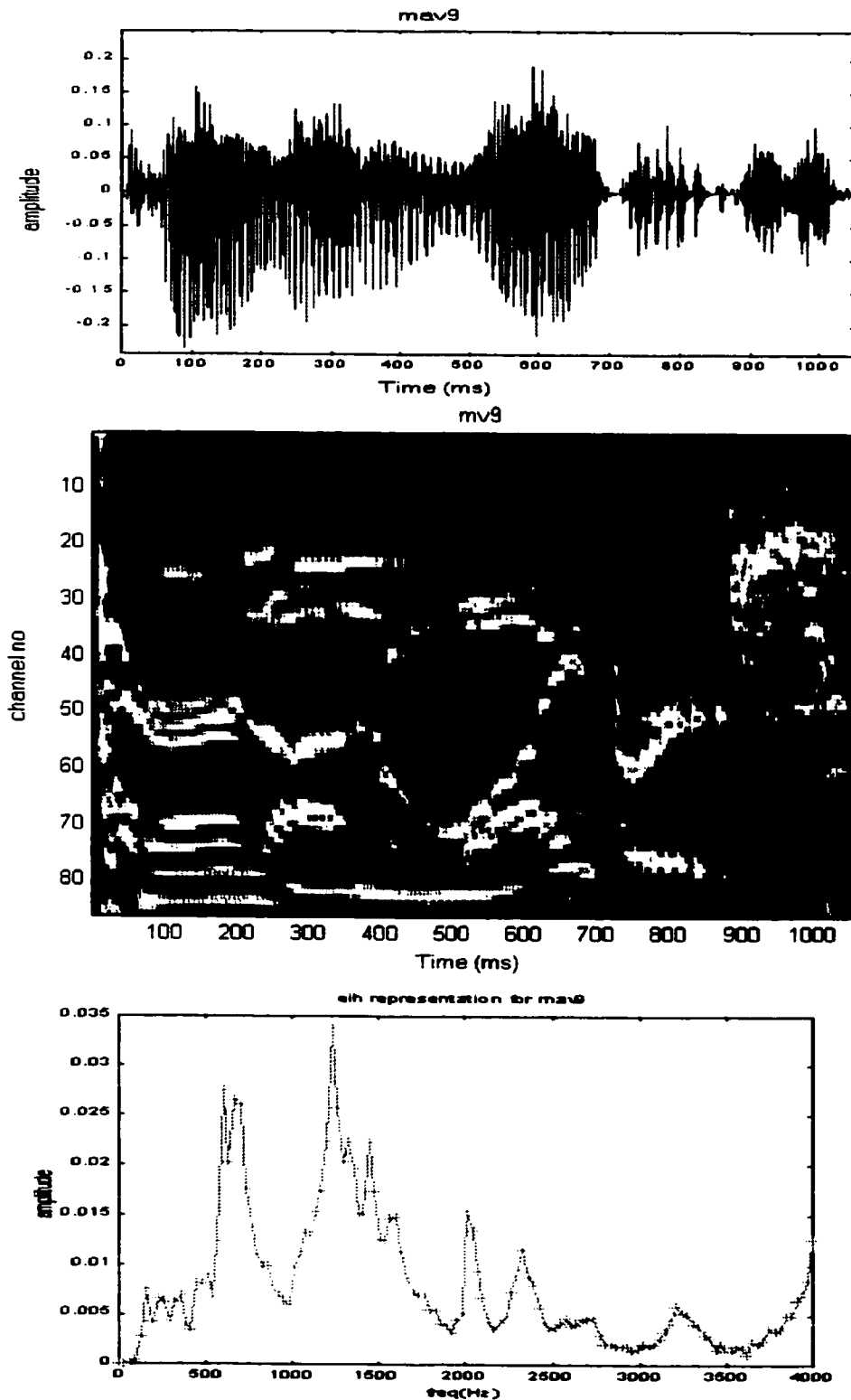


Fig 4.5(c) Sound Wave (upper), Cochleagram (middle), and EIH (below) of V9

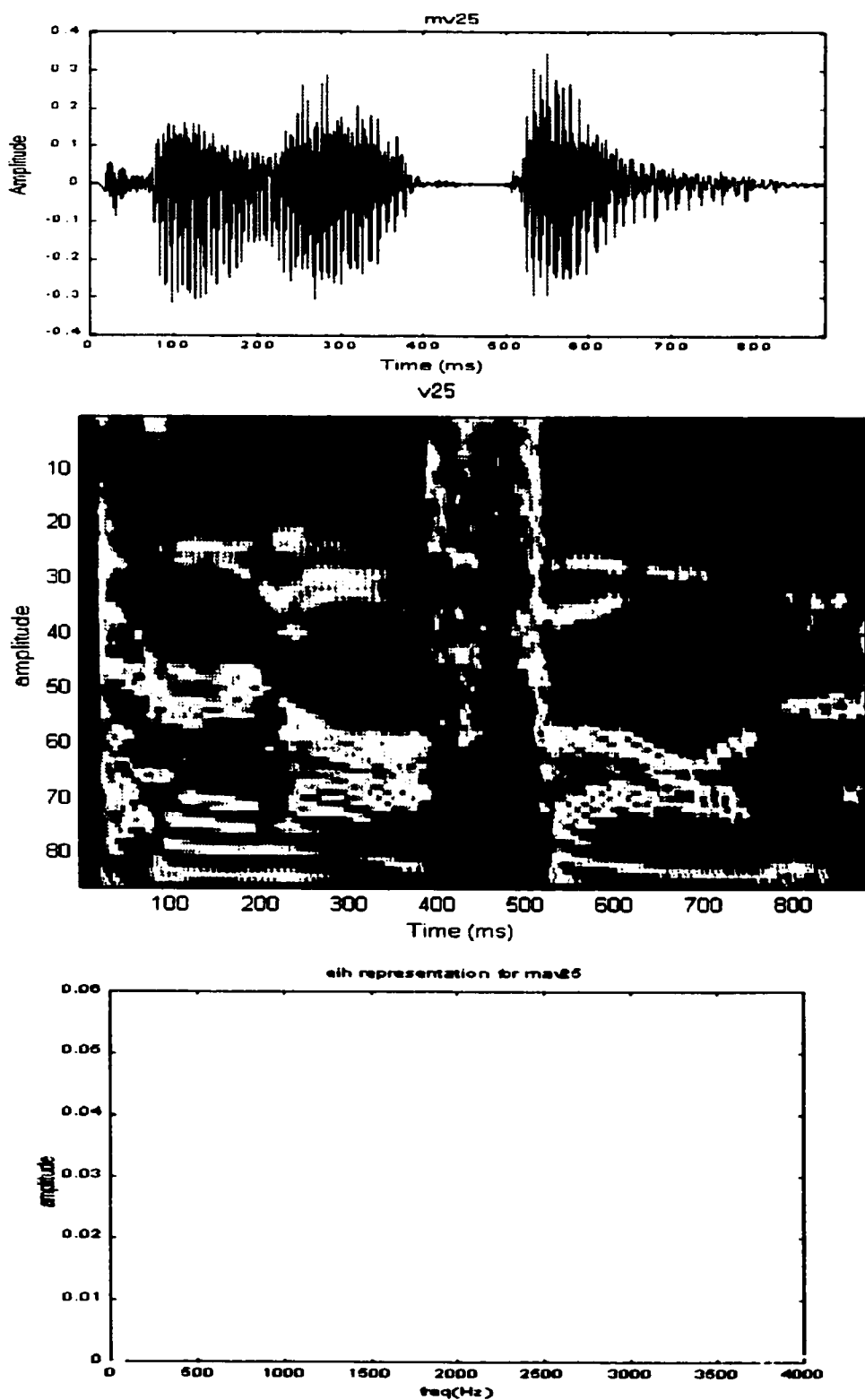


Fig 4.5(d) Sound Wave (upper), Cochleagram (middle), and EIH (below) of V25

Table 4.1 illustrates the 50 commands from 6 people recognition error rate.

Table 4.1 Recognition Error Rates for 50 commands

	32 thresholds, 128 linear bin allocation	16 thresholds, 128 linear bin allocation	32 thresholds, ERB bin allocation
v1	0	1	1
v2	1	1	2
v3	1	0	1
v4	1	1	0
v5	1	0	1
v6	0	1	0
v7	1	0	1
v8	1	1	0
v9	1	0	0
v10	0	1	0
v11	1	0	1
v12	0	2	1
v13	1	0	2
v14	1	2	1
v15	0	0	2
v16	0	1	0
v17	1	0	2
v16	1	1	1
v19	0	1	2
v20	1	1	1
v21	1	0	0
v22	0	1	1
v23	0	1	0
v24	0	1	0
v25	1	0	1
v26	1	1	1
v27	1	0	1
v28	1	0	0
v29	0	0	1
v30	2	1	0
v31	1	1	1
v32	0	1	0
v33	1	1	1
v34	0	1	0
v35	1	1	1

	32 thresholds, 128 linear bin allocation	16 thresholds, 128 linear bin allocation	32 thresholds, ERB bin allocation
v36	0	1	1
v37	1	1	1
v38	0	0	1
v39	0	1	0
v40	1	0	1
v41	0	1	0
v42	1	0	0
v43	0	1	1
v44	0	0	1
v45	1	0	1
v46	0	1	1
v47	0	0	1
v48	0	0	0
v49	0	0	1
v50	0	0	0
overall	27	32	37

Three points are noteworthy:

- The overall recognition rate (recognition rate = 273/300) is little more than some other recognition methods.
- On the average, linear bin allocation EIH performs better than the ERB allocation (error rate = 37/300) although the latter represents the human auditory system much closer. It is due to too much higher frequency components losing in ERB bin allocation.
- The average recognition rate is increased with the increase of the number of crossing level. This is because the more crossing level representation more accurately represents the features of the speech. But there is a tradeoff between the crossing level number and the computational load because the increasing of the crossing level number will increase the loop number in the EIH algorithms.

C h a p t e r 5

Conclusions and Future Works

Section 1 Introduction

Section 2 Cochlear Model

Section 3 Ensemble Interval Histogram

Section 4 Evaluation

Section 5 Future works

1. Introduction

In this thesis we demonstrate an auditory model on speech recognition related task. We first describe the cochlear model and then describe the EIH representation on speech recognition. To examine the capability of the model, we employ the LVQ algorithms to evaluate the recognition accuracy.

2. Cochlear Model

In this thesis a signal processing model of the cochlea is described at first. The model represents the mechanical filtering and neural transduction of the mammalian auditory system. This model describes the propagation of sound in the inner ear and the conversion of the acoustical energy into neural representations.

2.1 Filterbank

The major element in the model is cochlear filter bank. It is modeled as a cascaded of the second order filters which center frequency from higher to lower. Each filter stage in the cascaded is the combination of a notch filter and a resonant filter. The notch filter operates at successfully lower frequencies so the net effect is to gradually low pass filter the acoustic energy. The resonant filter picks out a small range of the traveling energy and models the basilar membrane motion.

2.2 Second Filter

The experiment shows that a notch in the frequency response approximately one octave below the center frequency. In order to model this effect of a second filter is followed each notch/resonant filter in the cascaded model. This filter vividly models this notch in the cochlear frequency response.

2.3 Half Wave Rectifier and Nonlinear Stage

The half-wave rectifier models the detection nonlinear stage of the hair cell. It provides a crude energy measure in the signal. The nonlinear stage in the model is used to attenuate the output of the half wave rectifier. Meanwhile, it simulates the ear adaptation to the signal amplitude. Third, the nonlinear stage also models the masking effects of the ear, i.e. the output of the given channel of the cochlear model not only depends on this channel itself but on its nearest neighbors.

2.4 Four-multiply Normalized Ladder Filter Implementation

In the thesis the second order filters are implemented using four-multiply normalized ladder filter form. As the filter transfer function is decided, choosing the best filter implementation strategy is one of the most difficult jobs to the filter designer. The main advantage of ladder form implementation is every node in the filter is normalized in the L_2 norm. This will result in absence of the overflow in the filters. Another advantage of this implementation is orthogonal tuning, i.e. in the second order cases the filter center frequency and its Q value are controlled by independent parameters.

2.5 Filter Optimization Using Genetic Algorithms

The objective of the thesis is to design the cochlear model to simulate the human auditory system. So we also employed the genetic algorithms to optimize the filter parameters so that the frequency response of the cochlear model is much closer to the experiment data than the conventional method. As shown in previous chapter the GA greatly increased the fitness of the model to the physiological data.

3. Ensemble Interval Histogram

In this thesis we also explored the ensemble interval histogram (EIH) representation in the speech recognition. This representation models the properties of the post auditory-nerve nuclei in the human auditory model. It is derived by collecting the level crossing interval from all the channels to form a frequency component representation of the acoustic signal. As such, it differs markedly from traditional speech representation methods and, consequently, exhibits quite different properties of the sound signal.

4. Evaluation

The extent of the accuracy of the recognition is evaluated in a quantitative manner by recognizing 50 commands. It shows the model performs well for the short command. This is probably due to the short commands consist of more distinctive EIH component than the long commands.

5. Future Works

In this model only one-dimensional wave propagation is modeled. The two or three dimensional ducts could describe the cochlear model more accurately, therefore , more accurate recognition will be obtained.

Another approach in the cochlear model is the active model. In this model a active system adjusts the Q values of the filter in the cochlear model in response to the filter output level, thereby providing a mechanism for the hypothesized out hair-cell behavior in changing the cochlear filter shape with signal level.

References:

1. Jont B Allen, *Magnitude and phase-frequency response to single tone in the auditory nerve* Journal of the Acoustical Society of America, vol 73, pages 2071-2092 1983
2. Jont B Allen, *How do human process and recognize speech?* IEEE transactions on speech and audio processing, vol 2, No. 4, pp 567-577 1994
3. Jont B Allen, *Cochlear modeling* IEEE ASSP magazine Jan. 1985 pp 3-19
4. Jont B Allen, and Stephen T. Neely, *Micromechanical model of the cochlear* Physics Today July 1992 pp40-47
5. T. Arsian and D.H. Horrocks, "A Genetic Algorithm for the Design of Finite Word Length Arbitrary Response Cascaded IIR Digital Filters", *Genetic Algorithms in Engineering System: Innovations and Applications*, Sept. 1995 No. 414, IEEE, 1995, pp276-281
6. Peter Dallos, *The active cochlear* The journal of neuroscience, Dec. 1992 pp4575-4585
7. K. De Jong, "Learning with the Genetic Algorithm: An Overview", *Machine Learning*, vol.3 pp. 121-137, Oct. 1988
8. B. Delgutte and N.Y.S. Kiang, *Speech Coding in the Auditory Nerve: I Vowel-like Sound*, Journal of Acoustic Society of American, Vol. 75, pp. 866-878, 1984
9. Oded Ghitza, *Auditory model and human performance in tasks related to speech coding and speech recognition*, IEEE transactions on speech and audio processing, vol 2, no 1 Jan 1992 pp 115-132
10. D.H. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, New York: Addison-Wesley, 1989
11. Julius L. Goldstein, *Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering* Hearing Research, 49 pp39-60 1990

12. Julius L. Goldstein, *Relations among compression, suppression, and compination tones in mechanical response of the basilar membrane: data and MBPNL model* Hearing Research, 89 pp 52-68 1995
13. William M. Hartmann, *Signals, sound, and sensation* Springer-Verlag New York, Inc ISBN 1-56396-283-7 1998
14. M. Haseyama, "A Method Quantizing Filter Coefficients with Genetic Algorithm and Simulated Addealing", *IEICE Trans, Fundamentals*, vol. E79-A, no. 8, pp1130-1134, Aug.1996
15. Simon Haykin, *Neural networks* Prentice Hall New York 1999 ISBN 0-13-273350-1
16. J.H. Holand, *Adaptation in Natural and Artificial System*, Ann Arbor, MI: The University of Michigan Press 1975.
17. James M. Kates, *Accurate tuning curves in a cochlear model* IEEE transactions on speech and audio processing, vol 1, No. 4, pp 453-462 Oct. 1993
18. James M. Kates, *A time-domain digital cochlear model* IEEE transactions on signal processing, vol 39, No. 12, pp 2573-2591 Dec. 1991
19. N.Y.S. Kiang and E.C. Moxon, (1974), *Tails of Tuning Curves of Auditory Nerve Fibers*, *Journal of Acoustic Society of American* vol. 55, no. 3, pp 620-630, 1974.
20. Teuvo Kohonen, *The self-organizing map* Proceedings of IEEE, vol. 78, no. 9, Sept. 1990, pp. 1464-1480
21. Kosir and J.H. Tasic, "Genetic Algorithms and Filtering", *Genetic Algorithms in Engineering System: Innovations and Applications*, Sept. 1995 No. 414, IEEE, 1995.
22. R.F. Lyon, (1982), *A Computational Model of Filtering, Detection, and Compression in the Cochlea*, *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing* (1982), vol 2, pp 1282-1285, 1982.

-
23. Dana C. Massie, *An engineering study of the four-multiply normalized ladder filter*, Journal of Audio Engineering Society, vol. 41, No. 7/8 July/August 1993
24. Ray Meddis and Michael Hewitt, *Virtual pitch and phase sensitivity of a computer model of the auditory periphery* Journal of the Acoustical Society of America, vol 89, pages 2866-2881 1991
25. B.C.J. Moore and B.R. Glasberg, *Suggested Formula for Calculating Auditory-filter Bandwidth and Excitation Patterns*, Journal of Acoustic Society of American, Vol. 74, pp. 750-735, 1983
26. Stephen T. Neely, *A model of cochlear mechanics with outer hair cell motility* Journal of the Acoustical Society of America, vol 94, pp 137-146 Jul-93 Nedzelnitsky, V. (1980). *Sound Pressure in the Basal Turn of the Cat Cochlear*, Journal of Acoustic Society of American vol. 68, no. 6, pp 1676-1689, 1980.
28. L.P. O'Mard and R. Meddis, *A computational Model of non-linear auditory frequency selectivity*
http://www.essex.ac.uk/psychology/hearinglab/projects/keele96/keele96b_toc.html
29. L.P. O'Mard, E.A. Lopez-Poveda, and Meddis, R. *Complex signal processing using non-linear auditory frequency selectivity*
<http://www.essex.ac.uk/psychology/hearinglab/projects/keele96/keele.html>
31. R.D. Patterson, J. Holdsworth, and M. Allerhand *Auditory Models as preprocessors for speech recognition* In: The Auditory Processing of Speech: From the auditory periphery to words, edited by M. E. H. Schouten (Mouton de Gruyter, Berlin) pp 67-83. 1992
32. Roy D. Patterson, Mike H Allerhand, and Christian Giguere, *Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform* Journal of the Acoustical Society of America, vol 98, pages 1890-1894 1995

-
33. James W. Pitton, Kuansan Wang, and Biing-hwang Juang, *Time-Frequency Analysis and Auditory Modeling for automatic recognition of speech* Proceeding of the IEEE vol 84, No. 9, pp 1199-1215 Sept. 1996
 34. John G Proakis and , Dimitris G Manolakis. *Digital signal processing: principles, algorithms, and application* Macmillan Publishing Company in USA ISBN 0-02-396815-X 1992
 35. John G, Proakis, Charles M. Rader, Fuyun Ling, and Chrysostomos L. Nikias, *Advanced Digital Signal Processing* Macmillan Publishing Company in USA ISBN 0-02-396841-9 1992
 36. Philip A. Regalia and Sanjit K. Mitra, *Tunable digital frequency response equalization filters* IEEE transactions on acoustic, speech and signal processing, vol assp-35 1, No. 1, pp 118-120 Jan. 1987
 37. Philip A. Regalia, Sanjit K. Mitra, and P. P. Vaidynathan *The digital all-pass filter: a versatile signal processing building block* proceeding of the IEEE vol 76, No. 1, pp 19-35 Jan. 1989
 38. Luis Robles, Mario A. Ruggero and Nola C. Rich, *Basilar membrane mechanics at the base of the chinchilla cochlear I. Input-output functions, tuning curves, and response phases*. Journal of the Acoustical Society of America, vol 80, pages 1364-1374 1986
 39. Mario A. Ruggero, Luis Robles and Nola C. Rich, *Basilar membrane mechanics at the base of the chinchilla cochlear II. Responses to low-frequency tones and relationship to microphonics and spike initiation in the VIII nerve*, Journal of the Acoustical Society of America, vol 80, pages 1375-1383 1986

41. M.B. Sachs and E.D. Young, *Encoding of steady-state Vowels in the Auditory Nerve: Representation in Terms of Discharge Rate*, Journal of Acoustic Society of American, Vol. 66, pp. 470-479, 1979
42. Malcolm Slaney, *Lyon's cochlear model*, Apple Technical Report#13 1988
43. Malcolm Slaney and Richard F. Lyon, *A Perceptual Pitch Detector*, ICASSP 90 pp357-360
44. Heming Zhao, *Design and DSP realization of cochlear filter* Acustica, vol 21, No. 4, pp 319-323 Jul-96

VITA AUCTORIS

NAME: Jie Zhang

PLACE OF BIRTH Shanghai, China

YEAR OF BIRTH 1962

EDUCATION Changfeng High School, Shanghai, China
1976 - 1980
Shanghai University of Science and Technology, Shanghai, China
1980 – 1985 B.Sc.
University of Windsor, Windsor, Ontario
1998 – 2000 M.A.Sc.