COMBINING MULTIBAND COMPRESSION AND CEFS IN HEARING AIDS

EFFICIENTLY COMBINING MULTIBAND COMPRESSION AND IMPROVED CONTRAST-ENHANCING FREQUENCY SHAPING IN HEARING AIDS

By

SHAHABUDDIN ANSARI, B.E.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Applied Science

McMaster University

Master of Applied Science (2005) McMaster University
(Electrical & Computer Engineering) Hamilton, Ontario

TITLE: **Efficiently Combining Multiband Compression and Improved Contrast-Enhancing Frequency Shaping in Hearing Aids**

AUTHOR: Shahabuddin Ansari, B.E. (NED Engineering University)

SUPERVISOR: Dr. Ian C. Bruce

NUMBER OF PAGES: xiv, 118

Abstract

Sensorineural hearing loss imparts two serious hearing deficits in hearing-impaired people: reduced dynamic range of hearing and reduced frequency selectivity. Psychophysically, these deficits render loss of speech audibility and speech intelligibility to a hearing-impaired person. Studies of an impaired cochlea in cats have shown that the hearing loss originates from damage to or complete loss of inner and outer hair cells. Neurophysiology of an impaired cochlea in cats shows that the tuning curves of the auditory nerve fibers become elevated and broadened. Amplification in hearing aids has been used to restore audibility in hearing-impaired people. Multiband compression has been commercially available in conventional hearing aids to compensate for the reduced dynamic range of hearing. However, little has been achieved to improve the intelligibility of speech in the hearing-impaired people. The aim of this thesis is to restore not only the speech audibility in a hearing-impaired person, but also to improve their speech intelligibility through some hearing-aid signal processing. The compensation technique used in this thesis for speech intelligibility is based on the hypothesis that a narrowband response of the auditory nerve fibers to speech signals ensure phonemic discriminability in the central nervous system.

Miller et al. [1999] have proposed contrast-enhancing frequency shaping (CEFS) to compensate for the broadband responses of the fibers to first and second formants (F1 and F2) of a speech stimulus. Bruce [2004] has shown that the multiband compression can be combined with CEFS without counteracting each other. In Bruce's algorithm, a multiband compressor is serially combined with a time-domain CEFS filter. The MICEFS algorithm, herein presented, is a combination of multiband compression and an improved version of CEFS implemented in the frequency domain. The frequency domain implementation of MICEFS has improved the time delay response of the algorithm by 10 ms as compared to series implementation proposed by Bruce. The total time delay of the MICEFS algorithm is 16 ms, which is still longer than the

standard time delay of 10 ms in hearing aids. The MICEFS algorithm was tested on a computational model of auditory periphery [Bruce et al., 2003] using a synthetic vowel and a synthetic sentence. The testing paradigm consisted of five conditions: 1) unmodified speech presented to a normal cochlea; 2) speech modified with half-gain rule presented to an impaired cochlea; 3) CEFS modified speech presented to the impaired cochlea; 4) speech modified with MICEFS presented to the impaired cochlea, and; 5) MICEFS-modified speech with some added noise in the formant estimation presented to an impaired cochlea. The spectral enhancement filter used in MICEFS has improved the synchrony capture of the fibers to the first three formants of a speech stimulus. MICEFS has also restored the correct tonotopic representation in the average discharge rate of the fibers at the first three formants of the speech.

Acknowledgements

First of all I am truly grateful to Almighty God, Who has enabled me to accomplish this work.

I would like to extend my thanks to Dr. Ian C. Bruce, whose compassionate supervision has been a source of motivation throughout my M.A.Sc. candidacy. I would like to thank Brent Edwards for providing the MATLAB code for multiband compression. I am thankful to Kamran Mustafa and Harjeet Bajaj for setting the groundwork of this thesis. I would like to express my gratitude to all my lab mates, especially, Jennifer Ko and Shamsul Arefeen Zilany, who diligently proofread the manuscript of this thesis. I am grateful to Cheryl Gies, Helen Jachna, Terry Greenlay and Cosmin Coroiu, who have always been there with smiling faces when I needed them. I would also like to thank Naomi Harte for the hardware implementation of this work. Last but not least, I would like to thank my wife, Aisha and my ever demanding children, Minhaj and Zhara. Without their sincere support, this accomplishment would be an unfulfilled dream.

I would like to dedicate this thesis to my mother, Tanweer Fatima, and to my late loving father, Badruddin Ansari.

[This work was supported by NSERC Discovery Grant 261736 and the Barber-Gennum Chair Endowment.]

Contents

List of Figures

List of Abbreviations

Abbreviation	Term
AGC	Automatic Gain Control
ANSI	American National Standards Institute
AVC	Automatic Volume Control
AN	Auditory Nerve
BF	Best Frequency
BM	Basilar Membrane
CEFS	Contrast-Enhancing Frequency Shaping
CVR	Consonant to Vowel Ratio
DSL	Desired Sensation Level
DFT	Discrete Fourier Transform
DTFT	Discrete Time Fourier Transform
ERB	Equivalent Rectangular Bandwidth
F1	First Formant
F2	Second Formant
F3	Third Formant
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
IFFT	Inverse Fast Fourier Transform
IHAFF	Independent Hearing Aid Fitting Forum
IHC	Inner Hair Cell
LGOB	Loudness Growth in half-Octave Bands
LTI	Linear Time Invariant
MICEFS	Multiband and Improved CEFS

Continued on next page ↪

↪ Continued from previous page.

Abbreviation	Term
NAL-R	National Acoustic Laboratories - Revised
NAL-RP	National Acoustic Laboratories - Revised and Profound
OHC	Outer Hair Cell
OLA	Overlap-and-Add
POGO	Prescription Of Gain and Output
PR	Power Ratio
PSTH	Poststimulus Time Histogram
REAG	Real-Ear Aided Gain
RMS	Root Mean Square
SPL	Sound Pressure Level
STFT	Short-Time Fourier Transform
VOT	Voiced Onset Time

List of Symbols

Symbol	Variable Represented
A	area
E	energy transferred by a sound wave through a given area
f	pitch or frequency of a waveform
f_c	center frequency of bandpass filters in a filterbank
$G50$	gain applied by a compressor to a input signal at 50 dB SPL
$G80$	gain applied by a compressor to a input signal at 80 dB SPL
H_{3FA}	average hearing threshold at 0.5, 1 and 2 kHz used in NAL-R
H_i	hearing threshold at i^{th} frequency
$\|H(\omega)\|$	absolute value of amplitude response of a filter
I	intensity of a sound wave
I_o	intensity of the faintest perceivable sound
IG_i	insertion gain at i^{th} frequency
k_i	additive fitting constant
P	spectral power of a sound signal
P_s	sound power
p	sound pressure
p_o	sound pressure of the weakest perceivable sound
PC	additive gain used in NAL-RP
$p(n)$	PSTH
Q_{10}	measure of sharpness of a frequency tuning curve
$R(m,k)$	synchronized rate at PSTH bin m and frequency bin k
T	time period of a waveform
v	velocity of particles in a medium

Continued on next page ↪

↪ Continued from previous page.

Symbol	Variable Represented
$w[n]$	window sequence
X	0.15 times H_{3FA} used in NAL-R
$X[k,m]$	STFT of a sequence with frequency index k and window advancement m
$\|X(\omega)\|$	absolute value of amplitude of a speech signal
$\hat{x}[n]$	resynthesized sequence using IFFT
$x_w[n]$	windowed speech sequence
z	impedance of a medium

Chapter 1

Introduction

Sensorineural hearing loss of cochlear origin is the most common type of hearing loss in the developed world. A person's dynamic range of hearing and perceptibility of speech would be reduced due to such hearing loss. Previous studies of auditory-nerve (AN) fibers in a damaged ear have shown a loss of sensitivity at the best frequency (BF) and a broadening of tuning curves [Kiang et al., 1976; Liberman & Mulroy, 1982; Salvi et al., 1982; Schmiedt et al., 1980]. As a result of the broadened tuning curves, the neural representation of a speech stimulus is degraded [Palmer & Moorjani, 1993; Miller et al., 1997; Schilling et al., 1998]. Normally, auditory nerve fibers synchronize their responses to the formants of a speech stimulus. This narrowband response of the fibers is called *'synchrony capture'*, which may help in perceiving different voiced phonemes. In an impaired cochlea the synchrony capture of the fibers is not observed. Instead, the fibers in the damaged cochlea exhibit a broadband response to a stimulus [Miller et al., 1997; Schilling et al., 1998].

Conventional hearing aids use multiband compression to compensate for the reduced dynamic range of hearing. Attempts have also been made to improve speech intelligibility in hearing aids by using spectral enhancement techniques. However, hearing-aid users complain that the sounds are fuzzy, unclear, muffled and distorted. Recently, a physiological approach has been sought in hearing aids to improve the speech intelligibility by hearing-impaired individuals [Sachs et al., 2002; Schilling

et al., 1998; Miller et al., 1999]. The motivation of this thesis is based on two independent studies presented by Miller et al. [1999] and Bruce [2004]. Bruce has shown that multiband compression and CEFS [Miller et al., 1999] can be combined without counteracting each other. The hearing-aid algorithm presented here is called MICEFS (Multiband compression and Improved CEFS), which efficiently combines multiband compression with an improved version of CEFS. MICEFS is tested using a model of auditory periphery [Bruce et al., 2003]. The results show an improvement in synchrony capture of the auditory fibers by the first three formants. Furthermore, an improved time delay response and average discharge rate of the fibers are also reported. MICEFS uses a formant tracker [Mustafa & Bruce, ress] for determining the frequency response of the time-varying spectral enhancement filter.

1.1 Current Hearing Aids and their Limitations

Hearing aids have been used to compensate for the losses associated with cochlear damage. The primary goal of most hearing aids is simply to restore audibility using frequency-selective amplification. Many hearing aids use linear amplification over most of their operating range, which is independent of the input level. A hearing aid can use various prescriptive rules to derive gain-frequency response from the hearing-loss profile, for example, half-gain rule [Lybarger, 1978], prescription of gain and output (POGO) [McCandless & Lyregaard, 1983], and a formula from National Acoustic Laboratories of Australia (NAL-R) [Byrne, 1986]. The problem with linear hearing aids is that they do not provide an effective way of dealing with the reduced dynamic range of hearing associated with cochlear damage. They equally amplify all sounds, which makes intense sounds over-amplified and uncomfortably loud.

A compression hearing aid is a nonlinear amplifier, which compresses the dynamic range of the input signal into a smaller dynamic range at the output. For people with much reduced dynamic ranges it may be difficult to maintain a volume control

setting that makes the weakest phonemes sufficiently audible without the most intense phonemes becoming too loud. A potential solution is to include a fast-acting or syllabic compressor that increases its gain during weak syllables or phonemes and decreases its gain during intense syllables or phonemes. Syllabic compression is usually employed in multiband compressors [Villchur, 1973]. Multiband compressors apply different amount of compression in different frequency bands, which depends on the hearing loss at that frequency and the level of the signal in that frequency region. Fast-acting multiband compression may increase audibility for people with a very reduced dynamic range of hearing, but it may also decrease intelligibility by altering the intensity relationship between phonemes [Plomp, 1994].

The reduced frequency selectivity is partially responsible for the reduced ability of people with cochlear hearing loss to understand speech. Linear amplification and multiband compression do not compensate for the effects of reduced frequency selectivity, although high-frequency emphasis can partially alleviate the upward spread of masking. Spectral enhancement techniques have been employed in hearing aids to improve frequency selectivity. In 1980, Boers modified the spectrum of a set of sentences by increasing the level differences between peaks and valleys, and then added some noise. The speech processing, however, reduced overall intelligibility. Summerfield et al. [1985] investigated the effect of narrowing the bandwidths of the formants of synthesized speech sounds. This narrowing of bandwidth resulted in sharper spectral peaks and greater peak-to-valley ratios. The latter form of signal processing improved the identification of consonants at the end of the syllables, but overall intelligibility did not improve. Simpson et al. [1990] described a method to increase differences in level between peaks and valleys in the spectrum of speech in noise. In the method, sampled segments of the signal were windowed, smoothed, spectrally enhanced, and then resynthesized using overlap-and-add technique [Allen, 1977]. The results from subjective testing showed small but reasonably consistent improvements in speech intelligibility for the processed speech, typically of 6 to 7%. Bunnell [1990] described

a method to enhance spectral contrasts. Contrasts were enhanced at middle frequencies of the stimulus by using a linear frequency scale. Small improvements were found in the identification of stop consonants (p, t, k, b, d, g) presented in quiet to subjects with sloping hearing losses. Baer and Moore [1993] used spectral enhancement similar to Simpson et al. [1990], but varied the amount and the bandwidth of the enhancement processing. Large amounts of enhancement decreased the intelligibility in noise. However, Franck et al. [1999] have shown that multiband compression tends to flatten the speech spectrum and consequently counteracts any benefits of spectral expansion schemes.

1.2 Physiological Basis of Hearing Aid Design

Conventional hearing aids do not restore hearing to normal, and their benefits are limited. Conventional hearing aids can partially compensate for the loss of sensitivity produced by cochlear damage, but they do not compensate for the loss of perception due to reduced frequency selectivity. Sachs and colleagues [2002] have proposed a biological basis for hearing aid design to potentially improve speech intelligibility in hearing impaired. In this approach, a signal processing scheme is employed in hearing aids to restore the narrowband neural response to speech stimulus in a damaged cochlea [Schilling et al., 1998; Miller et al., 1999].

The frequency selectivity of an AN fiber is often illustrated by the tuning curves, which shows the fiber's threshold as a function of frequency. The tuning curves of normal auditory fibers demonstrate low hearing thresholds and narrow tips [Evans & Wilson, 1973; Schmiedt et al., 1990]. The sharp tip of tuning curves helps fibers to exhibit phase locking to frequency components near the BF [Johnson & Swami, 1980]. In response to complex speech sounds, the fibers demonstrate compressive nonlinearities, in which fibers with BFs close to formants of the stimulus tend to respond to the formant frequencies exclusively. This so called synchrony capture of the fibers may be useful for vowel discrimination [Young & Sachs, 1979].

Fibers in a damaged cochlea show elevated and broadened tuning curves [Kiang et al., 1976; Robertson & Johnstone, 1979; Liberman & Mulroy, 1982; Robertson, 1982]. In response to vowel sounds, the fibers demonstrate phase locking to a broadband of the stimulus frequency components, and the capture phenomenon is less evident in impaired cochlea [Miller et al., 1997]. This deteriorates the temporal representation of the spectral shape in the auditory nerve fibers in response to speech sounds, which may be a potential cause for loss of speech intelligibility in hearing-impaired people.

Previous studies have investigated the correlation between the deteriorated discharge rate patterns in a damaged cochlea and degraded speech perception in a hearing-impaired person. Perceptual studies show that the internal representation of a vowel's spectrum in hearing impaired is flattened, where the spectral peaks associated with formants are reduced in contrast with the spectral valleys [Tasell et al., 1987; Leek & Summers, 1996]. This flattening of spectrum is assumed to arise from the broadening of the cochlear filters in the defective cochlea. The upward spread of masking in perceptual studies [Danaher & Pickett, 1975; Summers & Leek, 1997] also corresponds to the upward spread of F1 synchrony. The psychophysical interpretation of the upward spread of masking is the broadening of the perceptual filters especially at low frequencies [Glasberg & Moore, 1986], which is related to the broadened tuning curves at the lower frequencies [Liberman & Dodds, 984b]. The broadening nature of the temporal representation of complex stimuli clearly predicts that the impaired listeners should experience a difficulty in identifying the spectral peaks present in speech especially in background noise [Miller et al., 1997]. This prediction is consistent with previous studies where the performance of the subjects with sensorineural hearing loss degraded in the presence of competing sound sources [Glasberg & Moore, 1989; Smoorenburg, 1992].

1.3 Compensation Techniques for Broadened Cochlear Filters

From the previous discussion it is established that a damaged cochlea exhibits a broadband response to speech, and this loss in frequency resolution affects speech perception by hearing-impaired individuals. In response to vowel sounds, the synchrony is more broadband due to weak compressive nonlinearities of the fibers. The nonlinear suppression of the fibers, however, is not lost completely, and can be corrected by using some spectral enhancement schemes in hearing aids. The feasibility of designing a signal-processing scheme for hearing aids is achieved by using a model of the auditory periphery, which simulates the behavior of spiking neurons of normal and damaged cochleae [Bruce et al., 2003].

1.3.1 Frequency-Shaped Amplification

Schilling et al. [1998] have shown that commonly used frequency-shaped amplification schemes in hearing aids can deteriorate the neural responses to speech. Schilling and colleagues collected the neural responses in impaired-hearing cats to a vowel modified by amplification with frequency shaping based solely on the hearing loss. The scheme uses a highpass filter with passband gain determined by the half-gain rule [Lybarger, 1978]. For this particular hearing loss and vowel, the cutoff frequency of the filter happens to fall between F1 and F2 of the vowel as shown in Figure 1.1. Although the frequency-shaped amplification has improved the neural representation of the vowel by localizing the synchrony capture of BFs to F1, the high frequency gain also has amplified the harmonics in the trough between F1 and F2. This may cause a loss in sensitivity to subtle changes in the stimulus spectrum especially at F2 and deteriorates the vowel discrimination in a hearing-impaired person. The effects of reduced spectral contrast between formants and trough harmonics are more devastating in the presence of background noise [Leek & Summers, 1996].

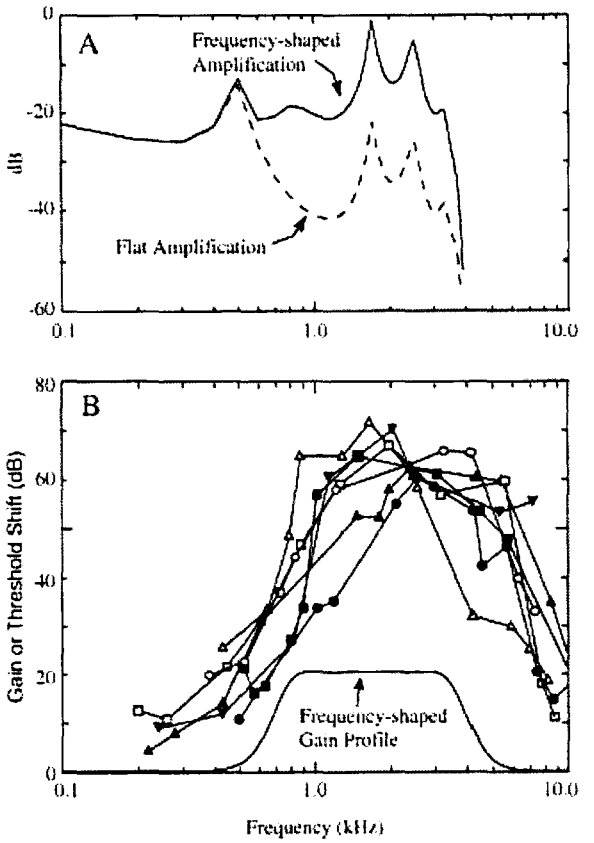

Figure 1.1: A: Spectrum of the vowel stimulus with flat amplification (dashed line) and with frequency shaping (solid line). B: Hearing loss profiles and gain-frequency response. Source: [Schilling et al., 1998]

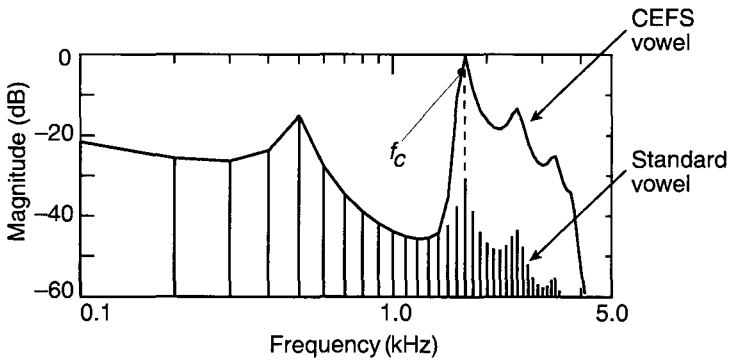

Figure 1.2: Line spectrum of vowel /ɛ/ with CEFS modified envelope. f_c is the cutoff frequency of the time-varying highpass filter used in CEFS, and is equal to F2 less 50 Hz. The time-varying highpass filter in CEFS ensures no harmonic amplification between F1 and F2. Source: [Miller et al., 1999]

1.3.2 Contrast-Enhancing Frequency Shaping (CEFS)

CEFS was proposed by Miller et al. [1999]. This scheme uses a time-varying highpass filter with cutoff frequency equal to F2 less 50 Hz of a vowel. Such a highpass filter applies gain only to F2 and higher frequencies without amplifying harmonics between F1 and F2 as shown in Figure 1.2. The results for a CEFS modified vowel show that the response to F1 has been localized, and the narrowband nature of F2 has been improved. When CEFS modified vowels differing only in F2 are presented to a defective cochlea, the fibers' responses show changes in rate and in cochlear place with changes in F2 frequency. However, in CEFS, the response to F3 is not restored because of the upward spread of synchrony to F2, and the amplification of harmonics in the trough between F2 and F3. Another potential issue with CEFS is that the tonotopic representation of the average discharge rate of the AN fibers at F2 of the vowel is not restored. The significance of fibers' synchrony and average discharge rate to speech intelligibility requires further investigation.

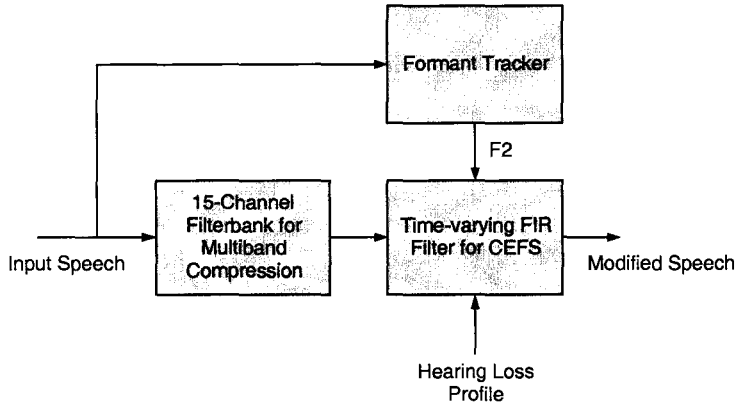

Figure 1.3: Schematic diagram of the series implementation of multiband compression and CEFS. The formant tracker estimates F2 of the input speech in real time, which with hearing-loss profile determines the frequency response of the FIR filter for CEFS. Modified from Bruce [2004].

1.4 Combined Multiband Compression and CEFS

From previous sections, it is obvious that fast-acting multiband compression can be used to match the reduced dynamic range of hearing with the dynamic range of speech, and CEFS seems promising for improving frequency selectivity in hearing-impaired people. Bruce [2004] has shown that multiband compression can be combined with CEFS without them counteracting each other. In his approach, he used a 15-channel filterbank for multiband compression implemented in the frequency domain, and a time-varying finite impulse response (FIR) filter for contrast enhancement implemented in the time domain. Figure 1.3 shows the schematic diagram of the series implementation of the two amplification schemes. The series implementation of the multiband compression and CEFS resulted in an overall time delay of 26 ms. This time delay is somewhat longer than required in hearing aids, which is known to create disturbances, and can diminish the speech perception of a hearing aid user [Stone & Moore, 1999, 2002, 2003].

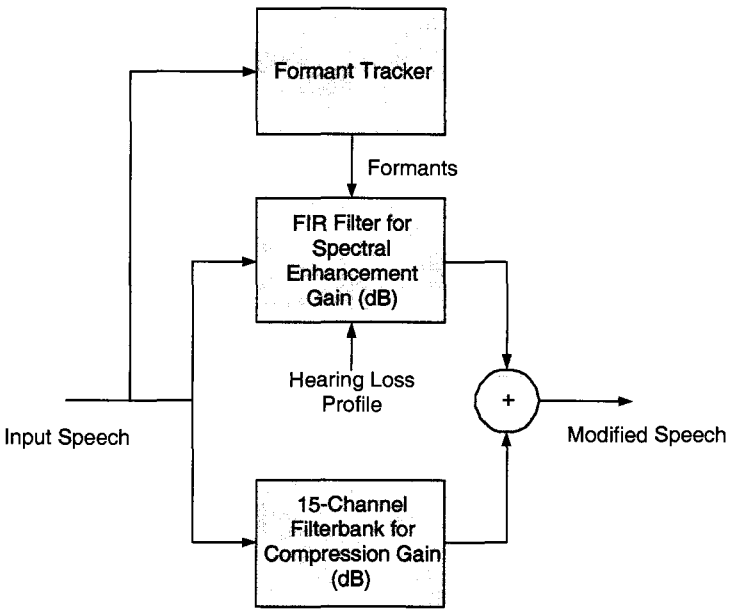

Figure 1.4: Schematic diagram showing the implementation of MICEFS. The spectral enhancement filter uses first four formants of the input speech and hearing-loss profile for its frequency response. The formants are estimated in real time by formant tracker.

1.5 Contribution of this Thesis

MICEFS is a hearing-aid algorithm developed for people suffering from sensorineural hearing losses of cochlear origin. MICEFS provides a hearing-aid solution for reduced dynamic range of hearing and reduced frequency selectivity in hearing impaired individuals by combining an improved version of CEFS with fast-acting multiband compression. The improved CEFS should restore the discharge rate pattern of the auditory nerve fibers in hearing-impaired individuals in response to voiced speech segments. This improves the frequency selectivity in response to voiced speech sounds, and thus, has the potential of increasing speech intelligibility by hearing-aid users. A schematic diagram in Figure 1.4 shows the implementation of MICEFS. The formant tracker [Mustafa & Bruce, In Press] determines the first four formants of the voiced part of the speech signal on frame by frame basis, which are used to design the time-varying FIR filter for spectral enhancement. The passband gain of the CEFS

filter is determined from the hearing-loss profile (audiogram) of a hearing-impaired person using a half-gain rule. The MICEFS has improved the neural representation of speech relative to original CEFS in the following ways:

- the upward spread of the synchrony to F2 is prevented by emphasizing F2 and F3 of the speech signal without emphasizing harmonics between F2 and F3;
- synchrony to F3 is restored by applying extra gain at F3 relative to the gain at F2 of the speech signal;
- the average discharge rate of the fibers in response to a speech signal is restored, which is almost the same as the normal discharge rate of the fibers;
- the overall time delay is 16 ms, which is 10 ms shorter than Bruce's implementation;

The restoration of the synchrony to the first three formants and fibers' discharge rate should improve the identification of voiced phonemes in hearing-impaired people. The major contributor of time delay is the size of the frame used in MICEFS, which is currently 512 samples. The reduction of this frame size to 256 ms may cause aliasing in time domain. The time delay of 16 ms is still longer than the desired maximal time delay in hearing aids, which is 10 ms. However, if the algorithm overwhelmingly improves speech perception, it may outweigh the disadvantages of the longer time delay in MICEFS. Final testing with human subjects is still required to determine its actual performance.

1.6 Thesis Layout

In Chapter 2 of this thesis, a brief background knowledge of relevant topics are presented. The chapter starts with some of the basic information about speech production and speech perception. It also discusses the physiology and pathophysiology

of a cochlea in sensorineural hearing loss, and the performance of some of the currently available hearing aids to address those cochlear impairments. At the end of the chapter, Bruce et al's model of auditory periphery [2003] is briefly described. Chapter 3 of this thesis discusses the signal processing methodology of the MICEFS algorithm. Chapter 4 and 5 present a detailed analysis of the results achieved from MICEFS and compared with CEFS and NAL-R prescription formula. Finally, the thesis concludes in chapter 6 with some of the accomplishments and potential issues with MICEFS.

1.7 Related Publications

Parts of this thesis have appeared in the following publication:

Ansari, S. U., Bajaj, H., Mustafa, K., and Bruce, I. C. (2004). "Time efficient contrast-enhancing frequency shaping and multiband compression in hearing aids". International Hearing Aid Conference (IHCON).

Bruce, I. C., Ansari, S. U., Bajaj, H. S., and Mustafa, K. (2004). "Multiband compression and contrast-enhancing frequency shaping in hearing aids," in Proceedings of the 2004 Annual Conference of the Canadian Acoustical Association, Ottawa, Ontario, October 2004.

Chapter 2

Background Literature

2.1 Physics of Sounds

Sound consists of mechanical waves, which are propagated as a disturbance in a medium. This disturbance arises as particles of the medium vibrate in the direction parallel to the direction of sound propagation. The sound waves produce compression and rarefactions in the medium, which are the high-pressure and low-pressure regions moving through the medium. This makes a sound wave a pressure wave. The time between two successive compressions or rarefactions is the time period T, and is measured in seconds. The pitch or frequency f of a sound is the number of vibrations of the medium particles per unit time, and is measured in Hertz (Hz). The time period and pitch of a sound wave are simply related as,

$$T = \frac{1}{f} \tag{2.1}$$

The intensity I of the sound wave is the amount of energy E transferred through a given area A of the medium per unit of time T. Since, energy per unit time is called power hence intensity is the sound power P_s per unit area and is measured in watt per meter squared. Mathematically,

$$I = \frac{E}{A.T} = \frac{P_s}{A} \tag{2.2}$$

Intensity can also be defined in terms of sound pressure p as,

$$I = \frac{p^2}{2z} = \frac{z.v^2}{2}, \tag{2.3}$$

where v is velocity of the vibrations of the particles in a medium with impedance z.

Sound pressure level (SPL) is a logarithmic measure of the root mean square (rms) value of sound intensity relative to the root mean square value of the intensity

of the faintest perceivable sound I_o, and is expressed in units of decibels. Decibel is a convenient scale in which equal increments in sound intensity roughly corresponds to equal increments in loudness. The decibel scale also represents large intensity changes with a narrow range of values. Analytically, SPL can be given as,

$$\text{SPL} = 10\log_{10}\left(\frac{\text{rms of I}}{\text{rms of I}_o}\right) = 20\log_{10}\left(\frac{\text{rms of p}}{\text{rms of p}_o}\right), \tag{2.4}$$

where p_o is the sound pressure of the weakest perceivable sound and its rms value is equal to 20 μPa (micropascal).

2.2 Production of Speech Sounds

There are three main groups of organs involved in the production of speech sounds: the lungs, the larynx and the vocal tract. The lungs provide the airflow to the larynx. The larynx modulates the airflow to provide either a periodic or a noisy source to the vocal tract. The modulation of the airflow is controlled by two masses of ligament called *vocal cords* or *vocal folds*. The slit-like orifice between the two folds is called the *glottis*. To create a periodic source, the vocal folds partially close the glottis to produce self-sustained oscillations. This state of the vocal cords is called the *voicing* state. The frequency of the oscillations is called the pitch or the fundamental frequency. *Unvoiced* sounds are generated by forming a constriction at some point along the vocal tract, and forcing air through the constriction to produce turbulence. The vocal tract then spectrally shapes the sources using oral, nasal, and pharyngeal cavities. To a periodic source, the vocal tract resonates at specific frequencies. These resonance frequencies are called *formants*. A third type of source can also be generated by constrictions in vocal tract, and is called an impulsive sound source. Impulsive source has more constriction than unvoiced sound source. The three sound sources combined with different vocal tract configurations produce *phonemes*.

2.3 Classification of Speech Sounds

A fundamental distinctive unit of a language is the phoneme. Syllables contain one or more phonemes, while words are formed with one or more syllables, concatenated to form phrases and sentences. Phonemes arise from a combination of vocal fold and vocal tract articulatory features. A broad classification of phonemes for English is shown in Figure 2.1.

Vowels belong to the largest phoneme group. The source of vowels is quasi-periodic puffs of airflow through the vocal folds vibrating at a certain fundamental frequency. Each vowel phoneme corresponds to a different vocal tract configuration or articulation. The particular shape of the vocal tract determines its resonances or

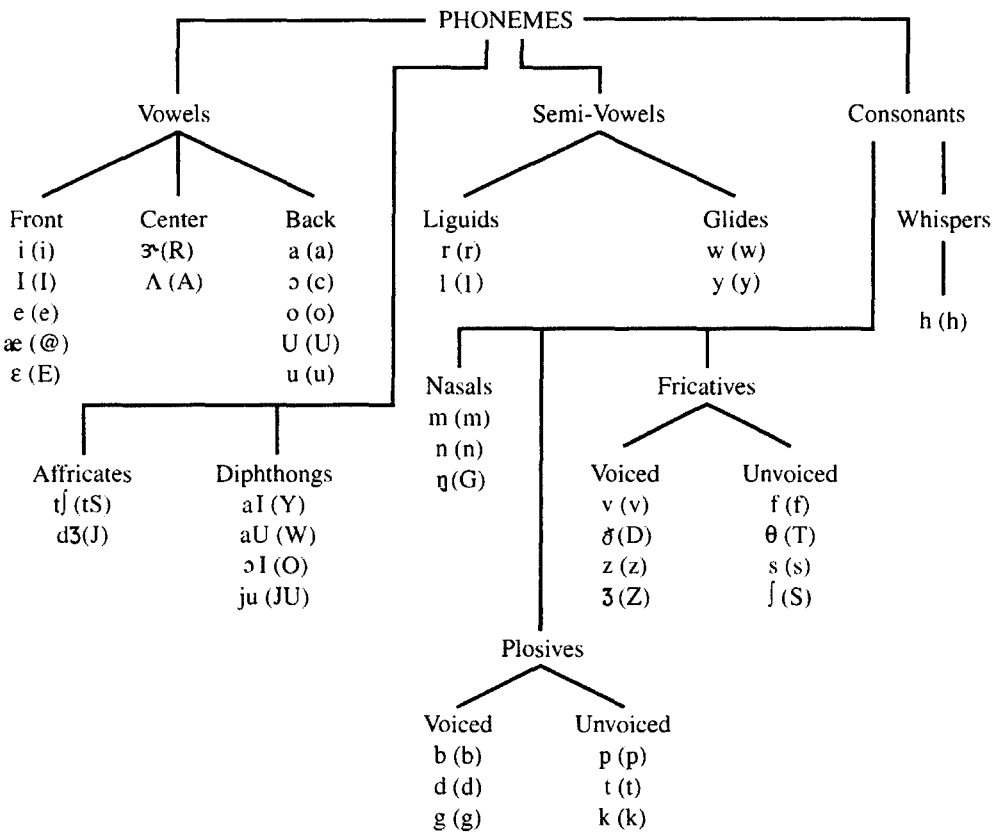

Figure 2.1: Phonemes in American English. Source: [Quatieri, 2002].

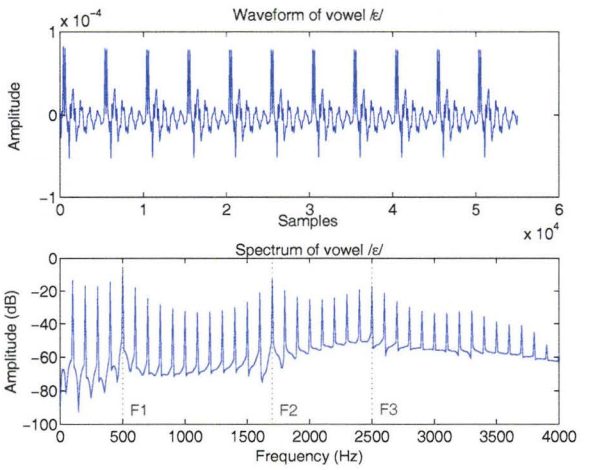

Figure 2.2: Waveform and spectrum of a vowel /ɛ/. The vertical dashed lines in the spectrum subplot show the first three formants of the vowel at 500, 1700 and 2500 Hz.

formant frequencies. Figure 2.2 shows time waveform and frequency spectrum of a vowel /ɛ/.

Consonants belong to the second largest group of phonemes. Consonants are classified into subgroups: nasals, fricatives, plosives, whispers, and affricates. *Nasals*, such as /m/ and /n/, are closest to vowels. The source in nasals is quasi-periodic puffs of airflow from vibrating vocal folds. The airflow is mainly through the nasal cavity with the oral tract being constricted. The spectrum of a nasal is dominated by the low resonance of the large volume of the nasal cavity. The waveform for both /m/ and /n/ are dominated by the low, wide-bandwidth F1 formant.

Fricatives are classified into two classes: voiced and unvoiced fricatives. In unvoiced fricatives, a noise source is generated by turbulent airflow through the constriction along the oral tract. Voiced fricatives have a similar noise source, however, the vocal folds vibrate simultaneously to generate periodic noisy airflow. *Whisper* is a consonant, which is similar in formation to unvoiced fricatives. In whispers, the turbulent flow is produced at the glottis rather than at a vocal tract constriction. An example of a whisper sound is /h/ as in "he".

Plosives can also be classified as voiced and unvoiced plosives. An impulsive source (burst) is used to excite the oral cavity in generating a voiced or unvoiced plosive phoneme. There is a little delay after the burst and the onset of the following vowel. This delay is called *voiced onset time* (VOT). In voiced plosive the vocal folds also vibrate. Although, the oral tract is completely constricted, there is a propagation of low-frequency vibrations through the walls of the throat. Such vibration before the burst is called a *voice bar*. Figure 2.3 shows the schematic diagram of the voiced

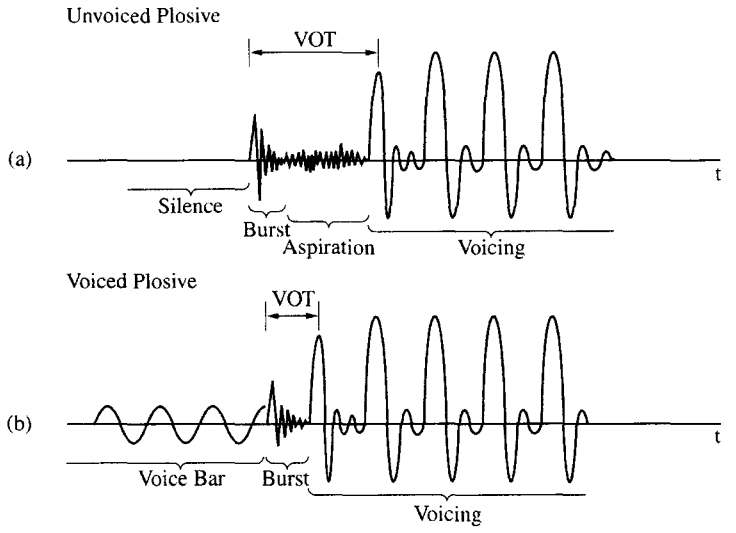

Figure 2.3: A schematic representation of (a) unvoiced and (b) voiced plosives. The voiced onset time is denoted by VOT. Source: [Quatieri, 2002].

and unvoiced plosives.

Some speech sounds are nonstationary where articulation changes from one phoneme to another. This time-varying vocal tract configuration results in transitional spectrum of speech sounds. *Dipthongs* are vowel-like sounds with vibrating vocal folds. However, in dipthongs, the vocal tract does not have a steady configuration, and it varies between two vowel configurations. Four dipthongs in the English language are: /Y/ as in "hide", /W/ as in "out", /O/ as in "boy", and /JU/ as in "new". Dipthongs are also characterized by transition between formants (especially F2).

Semi-Vowels are also vowel-like sounds with vibrating vocal folds. There are two types of semi-vowels: *glides* (/w/ as in "we" and /y/ as in "you") and *liquids* (/r/ as in "read" and /l/ as in "let"). Glides differ from dipthongs in two ways. In glides, the constriction of the oral tract is greater during the transition and the speed of the oral tract movement is faster that for dipthongs. The liquids differ from the glides in the formation of the constriction by the tongue; the tongue is shaped in such a way as to form side branches. The presence of these side branches can introduce anti-resonance.

Affricates are the counterparts of dipthongs. They consist of a combination of a plosive and a fricative, with the fricative consonant rapidly following a plosive consonant. The articulation of affricates is similar to that of a fricative with only one difference - they are preceded by a complete constriction of the oral cavity for the plosive.

2.4 Speech Perception

The acoustic properties of speech sounds act as perceptual cues for the auditory system. These perceptual cues are essential for phoneme discrimination by the auditory system.

In perception of vowels, the formant frequencies are a primary factor. Peterson and Barney [1952] have shown that a vowel can be identified by first two formants (F1 and F2). Pickett [1980], however, has shown that the higher formants also contribute to vowel identification. The formant location of a vowel is the function of the vocal tract length of a speaker. A listener has to normalize the formant location in vowel recognition by using relative formant spacings [Summerfield & Haggard, 1975].

The perception of consonants depends on many acoustic features of speech sound. The features include the formants of the consonants, formant transitions into the formants of the following vowel, the voicing or unvoicing of vocal folds, and the relative timing of the consonant and the onset of the following vowel. In the perception of plosives, the voicing and unvoicing can be identified by the VOT. The rate of a formant transition is also an important cue in perceiving consonants.

2.5 Anatomy of Human Ear

An ear can be broadly divided into three parts: outer ear, middle ear and inner ear. Each part of the ear performs a specific function to facilitate hearing. Figure 2.4 show the anatomy of a human ear. The anatomy and physiology of each part are briefly described below.

2.5.1 Outer Ear

The outer ear consists of a cartilaginous flange called the pinna, a cavity called the concha, an ear canal or external auditory meatus leading to the eardrum or tympanic membrane. The tube like structure of the outer ear helps to create resonances in order to influence the sound pressure at the tympanic membrane. Secondly, it helps in sound localization by providing the directionality cues in binaural hearing. The effect of resonance in the outer ear increases the sound pressure level from 2 to 7 kHz with a peak of about 15 to 20 dB at 2.5 kHz [Wiener & Ross, 1946; Shaw, 1974]. The outer ear is more sensitive to sounds originating in the horizontal plane at 45°. When the sound source is in the front of listener, the intensity and the timing differences of the sound at the two ears furnish the most important cues for locating a source to the left or right of the listener. For sounds generated in front of or behind the listener the information of localization comes from the filter properties of the pinna and concha of the outer ear.

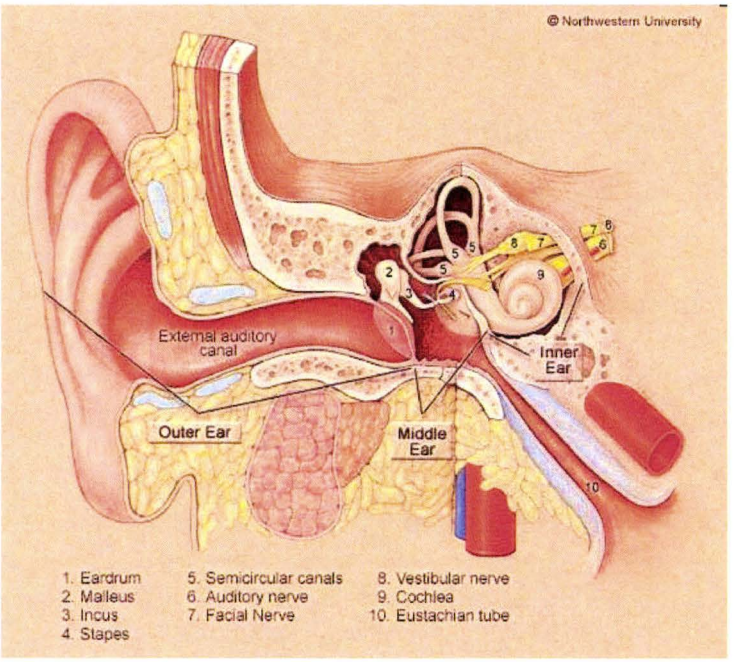

Figure 2.4: The anatomy of a human ear: the outer ear consists of the pinna, the concha and the external auditory meatus; the middle ear consists of three bones; malleus, incus and stapes; and the inner ear consists of the cochlea. Source: Webpage of Timothy C. Hain's, MD, Northwestern University.

2.5.2 Middle Ear

The middle ear is made of a bony structure called ossicles. The ossicles consist of three tiny bones: malleus, incus and stapes. The ossicles help to match the impedance of air in the outer ear to the higher impedance of the cochlear fluids. The impedance of the cochlear fluid is equal to that of seawater, i.e., $1.5 \times 10^6 \mathrm{N\,s/m^3}$. Another function of the ossicles is to apply force to the flexible oval window of the cochlea only. This ensures more fluid flow than it would if the pressure were also applied to the round window of the cochlea. The transmission of the sound energy is the greatest at 1 kHz where the sound pressure is about 30 dB greater than that at the tympanic membrane [Nedzelnitsky, 1980].

2.5.3 Inner Ear: Cochlea

The inner ear consists of a fluid filled coiled structure called the cochlea. The human cochlea is approximately 30 mm long. It comprises three chambers: *scala vestibuli*, *scala media* and *scala tympani* as shown in Figure 2.5. The scala vestibuli and the scala tympani are joined at the apex of the cochlea by an opening called the *helicotrema*. *Reissner's membrane* divides the scala vestibuli and the scala media. The *basilar membrane* divides the scala media and the scala tympani. The two outer scalae are filled with perilymph, an extracellular fluid. The scala media contains endolymph, an intracellular fluid with high concentration of K^+ and low concentration of Na^+ ions. The endolymph is at a high positive potential of approximately +80 mV, whereas the other scalae are at the potential of the surrounding bone. On the basilar membrane sits the *organ of Corti*, which constitutes the auditory transducer and the nerve endings. The nerve supply and the blood vessels enter the organ of Corti through the central cavity called the *modiolus*.

2.5.4 The Organ of Corti

The organ of Corti contains receptor cells called hair cells. The hair cells are of two types; *inner hair cells* (IHCs) and *outer hair cells*(OHCs). The IHCs are arranged in one row situated near the modiolar side of the cochlea. The OHCs are arranged in three to five rows with more rows at the apex. In total, there are about 15,000 of these cells in each human ear [Ulehlova et al., 1987] and 12,500 in cat [Schuknecht, 1960]. The hair cells have stereocilia at their apex. The stereocilia of the OHCs are shallowly embedded in the under-surface of the *tectorial membrane*, while the stereocilia of the IHCs loosely touch it. There are about 30,000 sensory neurons in man and 50,000 in cat innervating the cochlea. There are two types of auditory fibers; *afferent fibers*, which convey information to the central nervous system from the cochlea, and the *efferent or centrifugal fibers*, which receive information from the brain. About 90-95%

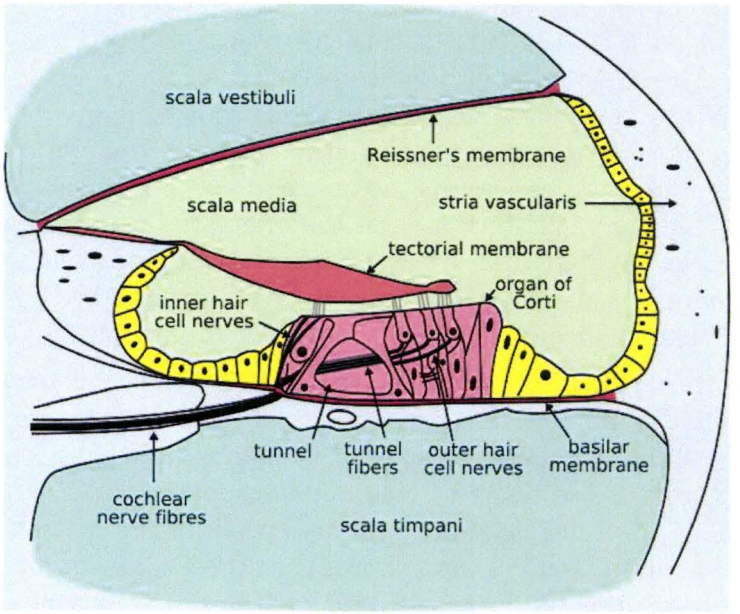

Figure 2.5: The anatomy of a human cochlea showing three chambers: scala vestibuli, scala media and scala tympani. The organ of Corti resides on the basilar membrane between the scala media and the scala tympani. The auditory nerve fibers innervate the inner and outer hair cells in the organ of Corti. Source: Wikipedia.

of the afferent fibers connect directly with the IHCs. The remaining 5-10% goes to OHCs. The efferent fibers mostly connect to the OHCs.

2.6 Physiology of the Cochlea

When sound enters the cochlea through the oval window it generates travelling waves in the basilar membrane (BM). The amplitude of the travelling wave increases to a peak at some point on the basilar membrane as it progresses toward the apex of the cochlea. Figure 2.6A demonstrates the transduction process that takes place in the cochlea. There is a relation between the place of the peak along the BM and the frequency of the incoming sound. Figure 2.6B shows the mapping of the fibers' best frequencies to the place along the basilar membrane. The motion in BM is more responsive to high frequencies towards the base and low frequencies towards the apex of the cochlea. Similarly, the hair cells in the organ of Corti are more sensitive to high frequencies at the base and low frequencies at the apex of the cochlea. This tonotopic behavior has also been experimentally confirmed in the auditory nerve fibers innervating the hair cells in cats [Liberman & Mulroy, 1982].

When the BM vibrates in response to a sound, it bends the apical stereocilia of the IHCs against the tectorial membrane. This bending of the stereocilia opens the ionic channels at the root of the stereocilia of the hair cells. The ionic current then triggers the auditory nerve fibers to create action potentials. The encoded speech by neurons is then transmitted to the auditory cortex in the central nervous system for perception. Previous studies have shown that the OHCs are mainly involved in controlling the motion of the BM, and act as automatic gain controllers (AGC) to amplify the softer sounds more than the intense sounds [Ruggero & Rich, 1991; Robles & Ruggero, 2001]. This behavior of the OHCs is mainly controlled by the efferent or centrifugal signals from brain [Pickles, 1988]. The nonlinear compression of the OHCs controls the motion of the BM at high levels of speech.

The fibers are characterized by their BFs. The BF of a fiber is the frequency of a tone, which triggers some response in the neurons at the lowest SPL called *threshold*. Figure 2.7 illustrates a typical tuning curve of a fiber with best frequency indicated by BF. Q_{10} is a measure of the sharpness of tuning, and is defined as the BF divided by the bandwidth of the tuning curve 10 dB above threshold. In response to a vowel sound, the normal auditory nerve fibers demonstrate strong synchrony to formant frequencies [Young & Sachs, 1979; Sinex & Geisler, 1983; Delgutte & Kiang, 1984; Palmer et al., 1986; Palmer, 1990]. The synchrony capture of the fibers is due to the their inherent compressive nonlinearities [Young & Sachs, 1979]. Figure 2.8 shows a method of analyzing auditory nerve responses to a vowel.

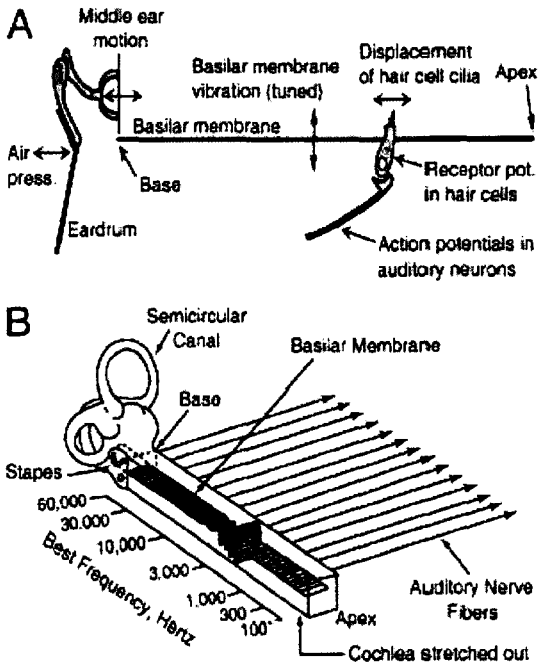

Figure 2.6: (A) BM vibrations displace the cilia of hair cells which transduce the vibration into electrical potentials that excite action potentials in auditory-nerve fibers. (B) The BM vibrations are tuned, so that energy at a given frequency causes a vibration which peaks at one point along the membrane. The scale at left shows the mapping of the best frequencies of the fibers into place along the BM for the cat cochlea. Source: [Sachs et al., 2002]

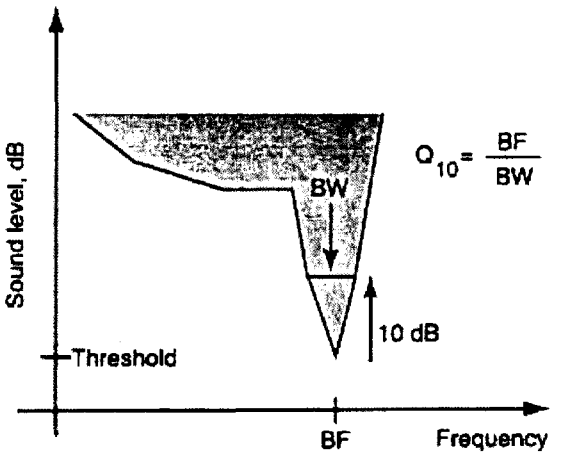

Figure 2.7: A tuning curve of a fiber with best frequency indicated by BF. A measure of sharpness of the tuning curve is given by Q_{10}, which is given as ratio of the BF to bandwidth of the tuning curve at 10 dB above threshold. Source: [Sachs et al., 2002]

2.7 Sensorineural Hearing Losses

Sensorineural hearing loss is a hearing threshold shift of 25 dB or more at 500, 1000, 2000 and 4000 Hz due to a defect in the cochlea or the auditory nerve whereby nervous impulses from the cochlea to the brain are attenuated. The sensorineural hearing loss of cochlear origin is more common and can be caused by acoustic trauma, disease, drugs or age (presbyacusis). Sensorineural hearing loss may be classified into two broad categories:

1. Temporary threshold shift resulting from an exposure to a noise. The loss is completely reversible and mostly affects the higher frequency regions of speech greater than 4 kHz.

2. Permanent threshold shift has occurred when hearing is not recovered completely after exposure to a noise.

2.7.1 Pathophysiology of the Cochlea

The impairment of the cochlea usually destroys the array of the stereocilia in some fashion; e.g., splaying, and uprooting of the stereocilia are common. Liberman and Dodds [1984b] had shown by careful experiments that in cases of damaged outer hair cells, the tuning curves were raised in threshold, and broadened in shape. In case of damaged inner hair cells, the tuning curves were raised in threshold, but the shapes

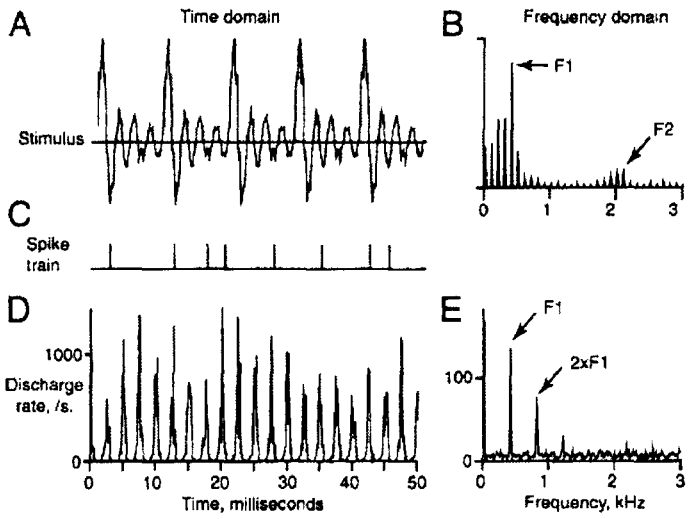

Figure 2.8: (A) Time-domain wave form of a 50 ms segment of the vowel /aɛ/ as in "bat". (B) Magnitude spectrum of the vowel. Spectral peaks corresponding to the first (F1) and second (F2) formants are indicated by the arrows. (C) Action-potential of an auditory-nerve fiber in response to one presentation of the vowel. (D) Mean instantaneous discharge rate vs time, obtained by averaging responses over many repetitions of the vowel. (E) Frequency spectrum of the mean instantaneous discharge rate. The fiber was tuned near the F1 frequency and consequently the instantaneous rate is phase locked to the F1 component of the vowel (the fourth harmonic). The higher harmonics of F1 in the response reflect distortion of the discharge rate response by rectification at zero rate. Source: [Sachs et al., 2002]

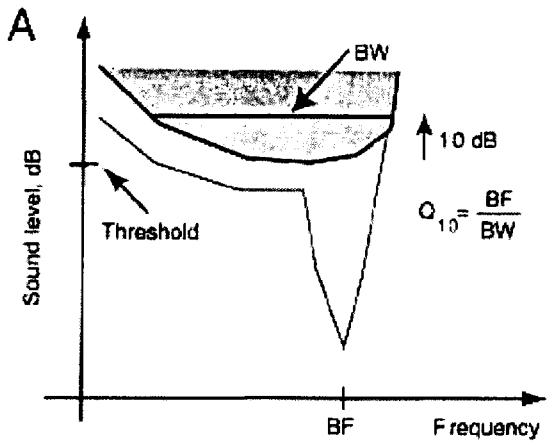

Figure 2.9: A tuning curve of an impaired fiber compared with a normal fiber with best frequency indicated by BF. The threshold of the impaired fiber has been elevated and the Q_{10} ratio has been deteriorated due to broadened bandwidth of the tuning curve. Source: [Sachs et al., 2002]

were normal. These results suggest that outer hair cells are involved in tuning of basilar membrane, and inner hair cells are involved in detecting the movement of the basilar membrane. In cases of severe cochlear damage both IHCs and OHCs are affected. Cochlear region suffering from complete IHC dysfuntion is called a "dead region".

The neural responses of auditory nerve fibers of a damaged cochlea to a tone have been shown to have an elevated threshold at BF and a broadening of tuning curves as shown in Figure 2.9 [Kiang et al., 1976; Robertson & Johnstone, 1979; Liberman & Mulroy, 1982; Robertson, 1982]. Two-tone suppression has also been diminished after cochlear impairment [Schmiedt et al., 1980; Salvi et al., 1982]. The elevation of the tuning curves increases the threshold of hearing without affecting the maximal comfortable loudness level, which results in a reduction in the dynamic range of hearing. The broadening of the tuning curves has devastating effects on the neural response to speech sounds. The discharge rate pattern shows a wide spread of neural activity in response to a tone and the tonotopic behavior of the fibers is not evident in the damaged cochlea. In response to a vowel, there is a broad response of the auditory fibers to F1, which masks the higher formant frequencies [Miller et al., 1997].

2.8 Compression in Hearing Aids

The major role of compression is to match the dynamic range of signals in the environment to the reduced dynamic range of a hearing-impaired person [Dillon, 2001]. A compressor is a nonlinear amplifier that automatically turns its gain down when an input signal rises to a certain level. A single channel compressor can be classified into compression limiters, syllabic compressors and automatic volume control (AVC) systems. The different compression schemes can be achieved by suitably modifying the static and dynamic characteristics of a compressor. *Static characteristics* of a compressor do not change with time. They include the following properties:

1. Compression threshold is the minimum input level required to activate the compression. This is also known as the *kneepoint*. Figure 2.10 shows the compression threshold at 40 dB SPL of the input level.

2. Static compression ratio defines the amount of output compression with increase in input levels. It is usually represented by a ratio of input levels in dB SPL of a pure tone required to produce 1 dB SPL of compressor output. Figure 2.10 shows compression ratio as the inverse of the slope of the input/output transfer function of a compressor.

3. Compression range is the range of input levels over which the compression is activated. Figure 2.10 shows the compression range starting from kneepoint to the maximum input level.

The *dynamic characteristics* of a compressor change with time, and refer to the temporal response of the system to the changes in input levels. The dynamic characteristics of a compressor comprise of the following properties:

1. Attack time is the time required by a compressor to decrease the output to within 3 dB (ANSI S3.22) of its final level when an input level changes from 55 to 90 dB (ANSI S3.22).

2. Release time is the time a compressor takes to increase the output level to within 4 dB (ANSI S3.22) of its final value when the input level changes from 90 to 55 dB (ANSI S3.22).

The static compression ratio is usually calculated for a pure tone. For speech signals, the actual or effective compression ratio is always less than the static ratio [Stone & Moore, 1992]. The static ratio is the measure of long-term compression whereas the effective ratio tells us the short-term compression. Although the dynamic range of the total signal level is compressed as described by the effective compression ratio, the dynamic range within a narrow frequency range is not compressed to the

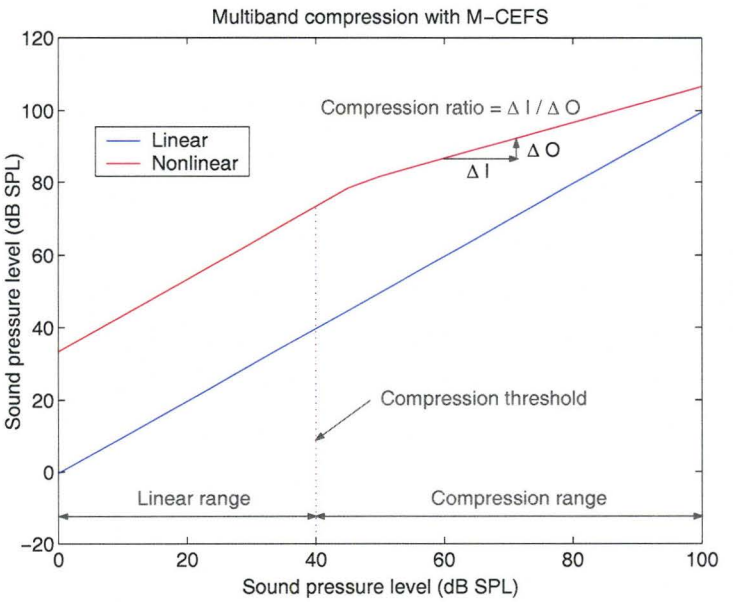

Figure 2.10: Plot showing compression of a speech signal. The compression threshold is 40 dB SPL. The compression ratio is defined as the inverse slope of the output versus input signal and is equal to 2:1. The region before compression threshold is called the *linear range*, and the region after compression threshold is called the *compression range*.

same degree as is the total broadband level [Verschuure et al., 1996]. This discrepancy occurs whenever the analysis bandwidth is less than the bandwidth of the signal passing through a compressor. The discrepancy is thus greatest for single channel compressors and least for multichannel compressors.

2.8.1 Compression Limiters

Compression limiting is a compression system with short attack time, high compression thresholds and large compression ratio (greater than 5:1). This is also known as a high level compressor [Walker & Dillon, 1982]. In such systems the amplification is linear at most of the input levels and average speech levels do not trigger compression. The compression limiters are similar in function to peak clippers. However, the limiters are superior to peak clippers, which introduce harmonic or inter-modulation distortion [Boothroyd et al., 1988; Braida et al., 1979; Dreschler et al., 1984; Dreschler, 988b; Preves, 1991; Walker & Dillon, 1982]. Compression limiting prevents discomfort and further damage to the aid user's residual hearing. The only disadvantage of this type of compression is that less gain is applied at input levels to the hearing-impaired person than with peak clipping.

2.8.2 Syllabic/Phonemic Compression

In syllabic or phonemic compression, the attack and release times are short, compression thresholds are low and the compression ratios are less than 5:1. In this compression scheme the levels of syllables of speech and phonemes within syllables are decreased. The release time ranges from 50 to 150 ms. It is vital that the release time should be less than the length of syllables, which is 200 to 300 ms [King & Martin, 1984]. People with sensorineural hearing loss and a reduced dynamic range of hearing typically use syllabic compression. Syllabic compression decreases the intensity fluctuations between vowels and consonants to allow for more of the speech signal to be audible without becoming too loud. A potential problem of fast acting compression is the relative change in the intensity relationship of the adjacent phonemes. The consonant to vowel ratio (CVR) is increased in syllabic compression to allow for consonant discrimination adjacent to a vowel. However, if a hearing-aid wearer makes use of this intensity relationship to identify different phonemes, altering relative intensities might affect his/her intelligibility of speech sounds [Dillon, 2001]. Another potential problem with syllabic compression is the decreased signal to noise ratio for noises occurring within the gaps of speech [Dillon, 2001].

2.8.3 Automatic Volume Control

Automatic volume control (AVC) is also called as *automatic gain control (AGC)* or *long-term compression*. The distinguishing feature of AVC is its long time constants. The release time is greater than 150 ms and could last for several seconds. The compression thresholds are usually low and compression ratios are greater than 5:1 [Walker & Dillon, 1982]. The output of the AVC is relatively constant in the presence of input fluctuations because of long release time and hence the aid user does not have to control the volume manually. Theoretically, the AVC could help in discriminating speech for hearing aid users who do not have severely reduced dynamic range of speech and they want to turn the volume of intense sounds down to a comfortable range [Braida et al., 1979; King & Martin, 1984; Walker & Dillon, 1982]. The potential problems with slow acting compressor are two fold. The first problem is that when a sudden loud sound appears, the compressor cannot decrease its gain in timely manner, which could make it uncomfortable for the aid wearer. A peak clipper or preferably compression limiter must then decrease the loud sound. The other problem is when a short soft speech follows a loud sound, the compressor would still be reducing gain and the soft sound may not be audible to the aid user. This can be alleviated by using release time no longer than needed.

2.8.4 Multiband Compression

Multiband compression allows different static and dynamic characteristics for different frequency bands. This type of compression is intended for those hearing-aid users who need different compression at different frequencies. It has been found that multichannel compressors improves signal to noise ratio and enhance speech perception in background noise. Multiband compressors generally use syllabic compression, however occasionally they use compression limiting [Bustamante & Braida, 1987] and AVC [Laurence et al., 1983; Moore et al., 1985]. The number of channels in multiband compression varies from two to sixteen.

There are some disadvantages of multiband compression in hearing aids. Multichannel compression decreases some of the essential differences between different phonemes. Because multiband compressors tend to decrease the height of spectral peaks and to raise the floor of spectral valleys, they partially flatten the spectral shapes. Since spectral peaks and valleys help in identifying speech sounds, spectral flattening makes it harder for the hearing-aid users to identify the place of articulation of consonants [Gennaro et al., 1986; Lippmann et al., 1981].

Table 2.3: Gain used in POGO formula

Frequency (Hz)	250	500	1000	2000	4000
k_i (dB)	-10	-5	0	0	0

2.9 Hearing-Aid Prescriptions

Hearing loss varies from person to person in degree, configuration and type. Hence there is a need for an amplification scheme that is suitable for a specific hearing-impaired person. Commercially available hearing aids can be broadly divided into two categories: linear hearing aids and nonlinear hearing aids. In hearing-aid prescriptions, the amplification is calculated from some measured characteristics of a hearing-impaired person. Knudsen and Jones [1935] proposed mirroring of the audiogram where gain at each frequency used in the hearing aid is equal to the threshold loss at the same frequency. Another method called the half-gain rule uses gain that is approximately half of the threshold loss [Lybarger, 1944]. Brief descriptions of some currently used prescription methods are given in the following paragraphs.

2.9.1 Linear Prescription Formula

Linear hearing aids have constant gain frequency response for all input levels. The linear hearing aids can be prescribed to improve the loudness comfort in quiet as well as noisy background [Dillon, 2001]. The *prescription of gain and output (POGO)* procedure uses the half-gain rule with additional attenuation at low frequencies [McCandless & Lyregaard, 1983]. The attenuation is intended to decrease the upward spread of masking from low-frequency ambient noise. The POGO formula is given as,

$$IG_i = 0.5H_i + k_i, \tag{2.5}$$

where IG_i is the insertion gain at i^{th} frequency, H_i is the hearing threshold at i^{th} frequency and k_i is the additive fitting constant at the i^{th} frequency given in Table 2.3.

For people with profound hearing loss (HL) POGO II was proposed [Schwartz et al., 1988]. The POGO II prescribes gain similar to that of POGO for losses less than 65 dB. For greater losses, the formula is modified as follows,

$$IG_i = 0.5H_i + k_i + 0.5(H_i - 65), \tag{2.6}$$

The *National Acoustic Laboratories of Australia (NAL)* formula was proposed to maximize speech intelligibility by hearing-aid users [Byrne & Tonisson, 1976]. The original NAL formula was revised for people with steeply sloping losses and became

Table 2.4: Gain used in NAL-R formula

Frequency (Hz)	250	500	1000	2000	3000	4000	6000
k_i (dB)	-17	-8	1	-1	-2	-2	-2

Table 2.5: Gain used in NAL-RP formula

H_{2kHz}	250	500	1000	2000	3000	4000	6000
≤ 90	0	0	0	0	0	0	0
95	4	3	0	-2	-2	-2	-2
100	6	4	0	-3	-3	-3	-3
105	8	5	0	-5	-5	-5	-5
110	11	7	0	-6	-6	-6	-6
115	13	8	0	-8	-8	-8	-8
120	15	9	0	-9	-9	-9	-9

known as NAL-R [Byrne, 1986]. The prescription formula for NAL-R is given as,

$$H_{3FA} = \frac{H_{500} + H_{1k} + H_{2k}}{3} \tag{2.7}$$

$$X = 0.15 H_{3FA} \tag{2.8}$$

$$IG_i = X + 0.31 H_i + k_i, \tag{2.9}$$

where k_i is given in Table 2.4.

For the severely and profoundly hearing impaired, NAL-RP was proposed [Byrne et al., 1991]. The formula of NAL-RP is given as,

$$X = 0.15 H_{3FA} + 0.2(H_{3FA} - 60) \tag{2.10}$$

$$IG_i = X + 0.31 H_i + k_i + PC, \tag{2.11}$$

where PC is an additive gain and is given in Table 2.5.

The *Desired Sensation Level (DSL)* prescription aims to provide the aid user an audible and comfortable signal in each frequency region [Seewald et al., 1993]. The DSL procedure uses desired sensation levels to calculate its target real-ear aided gain (REAG). At each frequency, REAG equals hearing threshold, plus desired sensation level, minus the short term maximum speech levels in the field for speech at an overall level of 70 dB SPL (see Dillon, 2001 for gain values).

2.9.2 Nonlinear Prescription Formula

Prescription formulas for nonlinear amplification have gain varying frequency response for several input levels. The purpose of nonlinear amplifier is to restore the normal loudness perception of the hearing impaired [Dillon, 2001].

Loudness growth in half-octave bands (LGOB) is an example of nonlinear amplification [Allen et al., 1990]. In this procedure, the patient categorizes the loudness of narrow bands of noise using a seven points loudness scale, which are compared to the same categories in normal-hearing people. For each input level, the gain needed to normalize the loudness is deduced.

Independent Hearing Aid Fitting Forum (IHAFF) is another example of nonlinear amplification [Valente & Vliet, 1990]. It uses loudness scaling to normalize loudness at each frequency. The particular loudness scaling procedure used is called the Contour Test.

Madsen Aurical method is another loudness scaling method with 7 loudness categories and is based on the procedure proposed by Kiessling et al. [1995].

ScalAdapt is an 11-point loudness scaling method [Kiessling et al., 1996]. It is performed while patient is wearing the hearing aid and clinician adjusts some characteristic of the hearing aid until the patient gives desired a loudness rating.

The *FIG6* procedure specifies the gain for medium and high-level input signals required to normalize loudness. It is based on the loudness data averaged across large number of people with similar degrees of threshold loss.

DSL[i/o] linear has linear input-output (I-O) curve for a wide range of input levels. Its compression ratio is constant within a wide compression region. The DSL[i/o] curvilinear has a curved I-O lines in the compression region. It is aimed at normalizing loudness.

NAL-NL1 attempts to maximize speech intelligibility, subject to the overall loudness of speech at any level being no more than that perceived by a normal hearing person (Dillon, unpublished data). It is a modification of the Speech Intelligibility Index method, in which allowance is made for the effects of hearing loss desensitization[1], and for the effects of listening at high SPLs.

2.10 Model of the Auditory Periphery

A computational model has been developed by Carney [1993] to describe the discharge patterns of auditory nerve fibers in cats. This model simulates almost all of the details of the signal processing in the cochlea. A time-varying narrowband filter determines the tuning of the BM. The OHC motion of the BM is handled by a

[1]The decreased ability of the impaired ear to extract information from signal even when it is audible.

nonlinear feedback path. The gain and the bandwidth of the time-varying filter are controlled by a feedback signal. The filter is sharp and linear at low intensity level, broad and compressive at moderate levels and broad and linear at high intensity level. These compressive nonlinearities of the BM are well documented in experimental animal studies. The Carney model does not include wideband nonlinearities of the BM, which is important in simulating speech data. Zhang et al. [2000] modified Carney's model to include the wideband nonlinearities and consequently predict the two-tone suppression. This is achieved by adding a wideband filter in the feedback path of the BM.

Bruce et al. [2003] further modified the model proposed by Zhang et al. by adding a middle ear filter and introducing a scaling function in the OHC and IHC section to describe the impairment of OHCs and IHCs. A modification in the control path is also made to improve the filter dynamics. With damaged OHCs the feedback path of the BM motion is also impaired, and the tuning curves are broadened and elevated. To model this impairment of the OHCs, the feedback path is scaled by a constant that varies from 0 to 1. For a normal cochlea the constant is set to 1. For a damaged cochlea the constant is set between 1 and 0, which means that the BM filter has lower gain and broader bandwidth at low intensity and less compression at moderate levels. Figure 2.11 shows the model schematic. The addition of the middle ear section has improved the model thresholds so that they now match the actual experimental data. The IHC impairment can be achieved by decreasing the slope of the function that relates BM vibration to IHC potential. A simulated audiogram for a hearing-impaired cat or person is shown in 2.12.

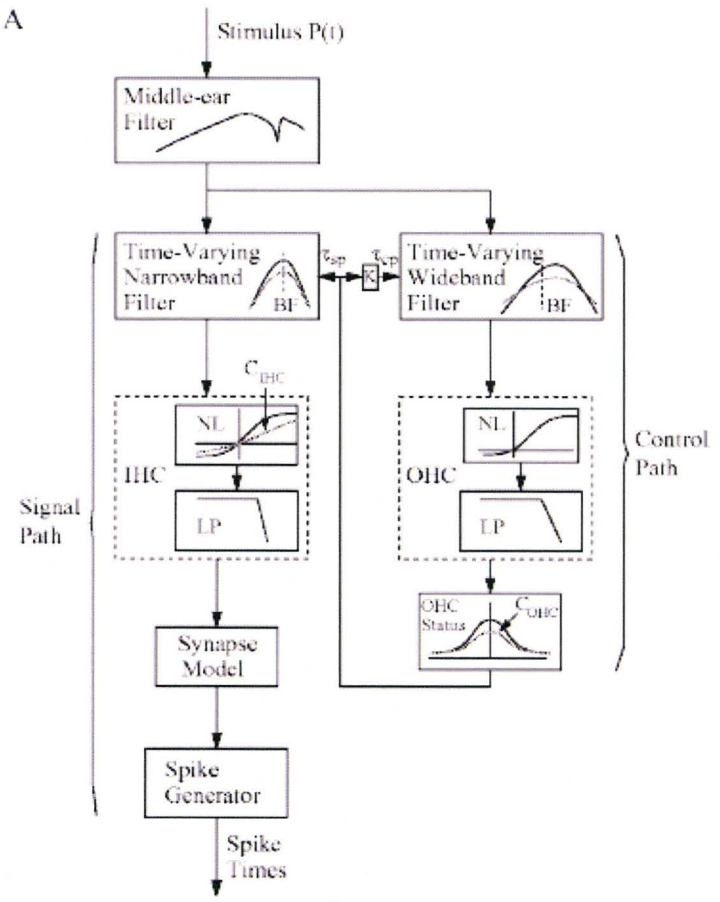

Figure 2.11: The auditory periphery model of the middle and inner ear. The inner ear shows IHCs in the forward path and OHCs in the feedback loop controlling the motion of the basilar membrane. The output of the model is the train of spikes simulating firing neurons. Source: [Bruce et al., 2003]

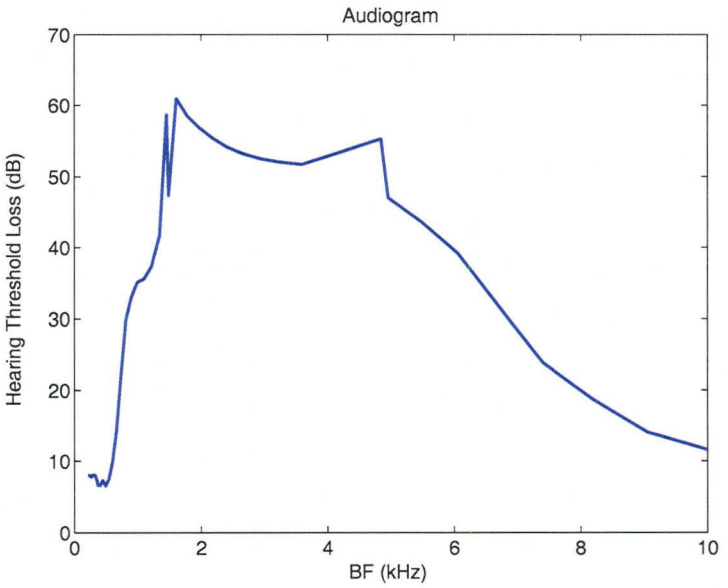

Figure 2.12: Audiogram of a hearing-impaired person simulated using model of auditory periphery [Bruce et al., 2003]. The audiogram shows moderate slopping high-frequency hearing loss of about 40 to 60 dB.

Chapter 3

Signal Processing Methodology

The MICEFS algorithm was implemented in MATLAB. Figure 3.1 shows the schematic approach adopted in developing the algorithm. Two types of speech signals, a vowel and a speech sentence, were used in the development of the algorithm. MICEFS uses formant information acquired from the formant tracker [Mustafa & Bruce, In Press] and the hearing-loss profile of a hearing-impaired individual to calculate the frequency response of the time-varying spectral-enhancement filter.

3.1 Speech Signals

One of the speech signals used in the development of the MICEFS algorithm is a vowel /ɛ/[1]. The original vowel is 110 ms long, and is normalized to 0 dB SPL as shown in the time domain representation in Figure 3.2. The vowel sequence is filtered with the human external ear transfer function [Wiener & Ross, 1946]. The first five formant frequencies of the vowel are 0.5, 1.7, 2.5, 3.3 and 3.7 kHz. A synthetic sentence used in MICEFS is "Five women played basketball"[2], which is shown in Figure 3.3. The test sentence is phonetically rich with a variety of formant trajectories. The presentation level for the sentence is given in terms of the SPL for the highest-amplitude phoneme, the 'a' in the word "basket" of the sentence. The duration of the original sentence is 1.6 s. In MICEFS, the speech signals (the vowel and the sentence) are sampled at 16 kHz. This sampling rate is sufficient to capture most of the spectral energy of the speech.

[1]The vowel /ϵ/ is extracted from the word "besh", synthesized by Klatt sythesizer [1980]
[2]The sentence was provided by R. McGowan of Sensimetrics Corp, Somerville, MA

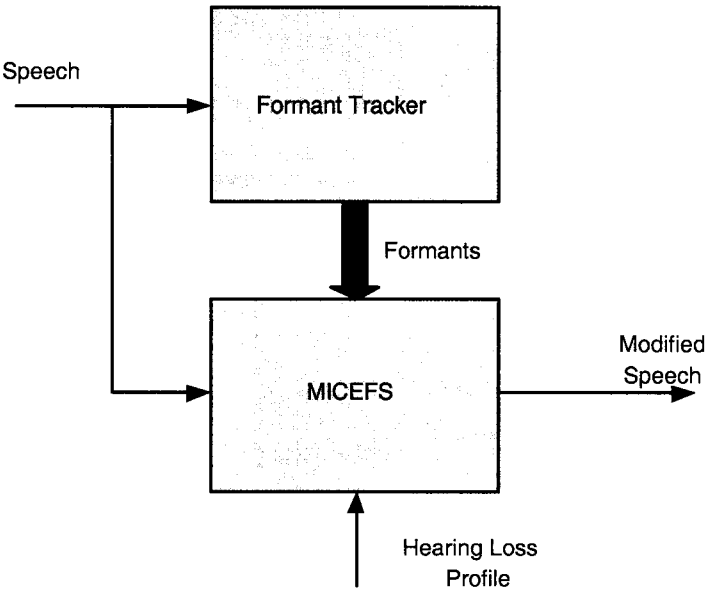

Figure 3.1: Schematic diagram of a hearing aid algorithm using the MICEFS algorithm. The formant tracker estimates the first four formants of a speech signal. The MICEFS algorithm uses the formant estimates and the hearing-loss profile to determine the frequency response of the time-varying spectral-enhancement filter.

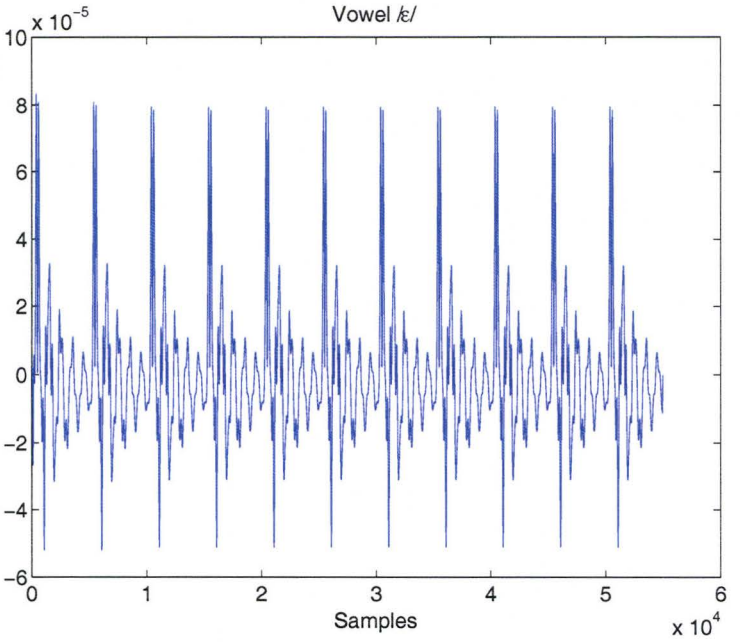

Figure 3.2: Time domain vowel /ε/ used in the development of MICEFS. The vowel is quasi-periodic with a pitch period equal to 10 ms. The vowel is 110 ms long, and sampled at 500 kHz.

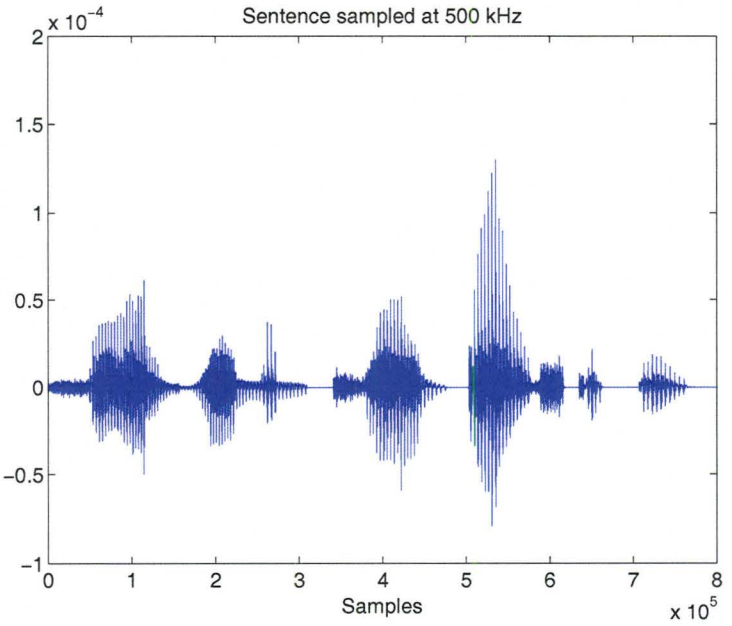

Figure 3.3: Time domain sentence "Five women played basket ball" used in the development of the MICEFS algorithm. The sentence is 1.6 seconds long with a sampling frequency of 500 kHz. The 'a' in basket is the most intensive component of the sentence.

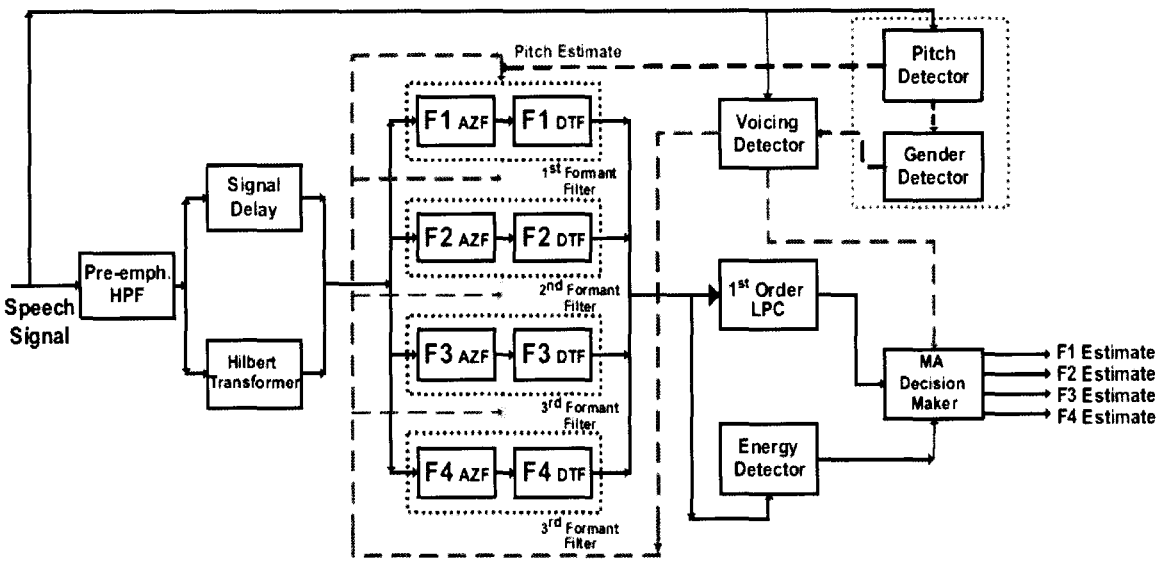

Figure 3.4: Block diagram of the formant tracker. The formant tracker relies on an adaptive filterbank to separate each formant frequency region prior to spectral estimation. Source: [Mustafa & Bruce, ress]

3.2 Formant Tracker

The formant tracker [Mustafa & Bruce, In Press] estimates the first four formants of the voiced speech using a set of time-varying adaptive filters. Figure 3.4 shows the schematic diagram of the formant tracker. The pre-filtering used in the formant tracker is included to limit the spectral region of the formant estimation and to minimize the effects of the neighboring formants and noise on the estimation. The formant tracker uses an adaptive voicing detector to determine when to track the formants. A pitch based gender detector in the tracker makes it robust in speaker variability, i.e., for both male and female speakers. It also uses an adaptive energy detector to detect the changes in overall speech energy over time, which makes it suitable for use with continuous speech. The energy detector makes the formant tracker robust in periods of silence, in background noise (additive white Gaussian noise), and in the presence of a competing background speaker. The formant tracker samples the input speech at 8 kHz, which is in contrast to the 16 kHz sampling rate used in MICEFS. To avoid a mismatch in time delay, a linear interpolation for the formant frequencies is used in MICEFS.

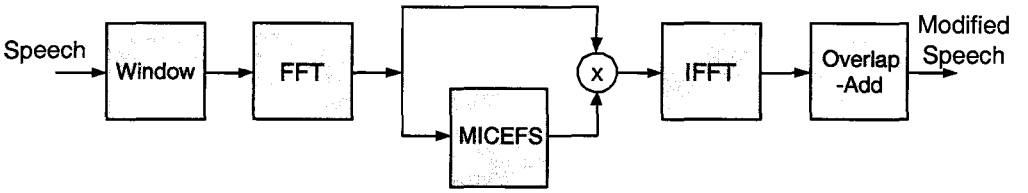

Figure 3.5: The implementation of MICEFS filtering. The MICEFS algorithm calculates the spectral gain (compressed and enhanced) in each frame and then applies it to the spectrum of the speech signal.

3.3 The MICEFS Algorithm

The MICEFS algorithm is implemented in the frequency domain by using an FFT-based 15-channel filterbank and a time-varying FIR filter. Figure 3.5 shows the schematic diagram of the frequency domain implementation of MICEFS. The window function is used to segment the speech signal into small frames. The FFT block estimates the spectral energy of the speech signal in each frame. The MICEFS algorithm evaluates the compression gain and the spectral enhancement gain for each speech frame. The gain just estimated by the MICEFS block is then applied to the input spectrum. Finally, time-domain modified speech is obtained by taking the inverse Fourier transform (IFFT) of each frame, and adding them together by using the overlap-and-add (OLA) method.

3.3.1 Analysis Window

A Hanning window is used in MICEFS for the speech analysis as shown in Figure 3.6. Although a rectangular window is relatively simple and provides relatively better frequency resolution, it has poorer side lobe spectral energy leakage. The peak side lobe in Hanning window is 32 dB down the main lobe, which is about 20 dB lower than the peak side lobe of a rectangular window. The width of the main lobe contributes to the frequency resolution of the analysis. The width of the main lobe in the Hanning window is $8\pi/(length\ of\ window)$, which is two times the width of the main lobe of the rectangular window. The Hanning window tapers down to zero, which avoids edge effect distortions with successive frames. In MICEFS, the duration of the window is 8 ms, which results in a frequency resolution of 125 Hz. Mathematically, the Hanning window is defined as a raised cosine function, i.e.,

$$w[n] = \begin{cases} 0.5 - 0.5\cos(2\pi n/M) & 0 \leq n \leq M \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

The linear convolution of finite data with finite impulse response of a linear time-invariant (LTI) system can be computed from circular convolution by multiplying the

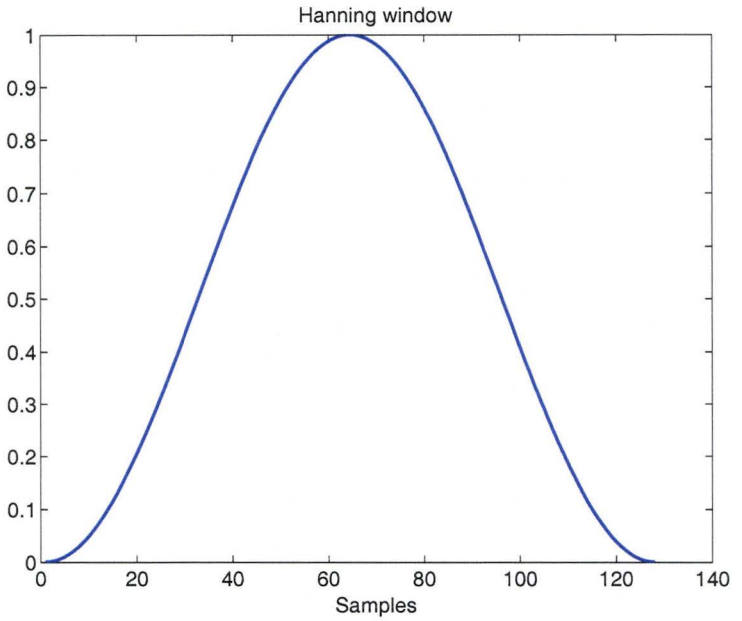

Figure 3.6: Time domain Hanning Window used in STFT. The duration of the window is 8 ms (128 samples sampled at 16 kHz).

DFTs of the two signals. For a filter length greater than 60 samples, it is faster to use the fast Fourier transform (FFT) than direct convolution. The number of operations required in FFT is $O(N \log_2 N)$, which is better than $O(N^2)$ operations required in time domain convolution. This framework also allows for filtering where the filter may be time dependent. The only consideration in frequency domain linear filtering is that the length of the DFTs should be equal to or greater than the length of the corresponding convolution sum. Zero padding is employed if data samples are less than the required length for frequency-domain linear filtering. The zero padding gives a better approximation of the FFT to the discrete time Fourier transform (DTFT) without improving the frequency resolution determined by the window function. For the efficient implementation of FFT, the length of the frame should be an integer power of 2.

In MICEFS, 128 data samples are zero padded with 384 zeros to form a 512-sample long frame. Half of the 384 zeros are placed at the beginning of the windowed data and the other half at the end of each frame. This data format avoids any phase distortion resulting from windowing. The window is then advanced 64 samples (half of the data) of the speech signal for the next frame. This 50% of data overlap provides adequate amplitude accuracy for the short-time frequency analysis of the speech signal. If the ultimate goal of the frequency domain analysis is to resynthesize the signal, then the window advancement S must be such that the overlapped window functions sum to

a constant c over all n, i.e.,

$$\sum_m w(n - mS) = c \tag{3.13}$$

3.3.2 Short-Time Fourier Transform (STFT)

A speech signal is quasi-stationary or short-time stationary. To perform Fourier analysis on the speech signal, we need to multiply the signal by a window time function. This gives us a time localized frequency spectrum of the speech. In MICEFS, a moving Hanning window is used for STFT. An example of the moving Hanning window is shown in Figure 3.7. The window length plays an important role in time-frequency resolution in the data analysis. The longer window results in narrowband representation of the data with poor time resolution. The shorter window represents better time domain resolution with broadband resolution in frequency domain. For each new frame the window is advanced 64 samples, which gives 50% speech data overlap. Windowing of N samples of the speech sequence $x_w[n]$ starting at N_o using a window sequence $w[n]$ is given in Eqn. 3.14.

$$x_w[n] = \begin{cases} x[N_o + n]w[n] & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

The STFT of the speech sequence is then evaluated by applying the discrete Fourier transform (DFT) to the windowed signal. Mathematically, the STFT can be given as,

$$X[k, m] = \sum_{n=0}^{N-1} x[n]w[n - m]e^{-2\pi nk/N}, \qquad 0 \leq k \leq N-1 \tag{3.15}$$

where k is the frequency index and m is the window advancement for each new frame.

3.3.3 Gain Calculation

The compression in MICEFS is realized by using a 15-channel filterbank. For spectral enhancement, an FIR filer with arbitrary gain is used. Figure 3.8 shows the schematic diagram of the gain calculation. The combined gain is then applied to the original speech spectrum in each frame.

Multiband Compression

A 15-channel filterbank of bandpass filters, as shown in Figure 3.9, is used for multiband compression. The bandwidth of each bandpass filter in the filterbank is

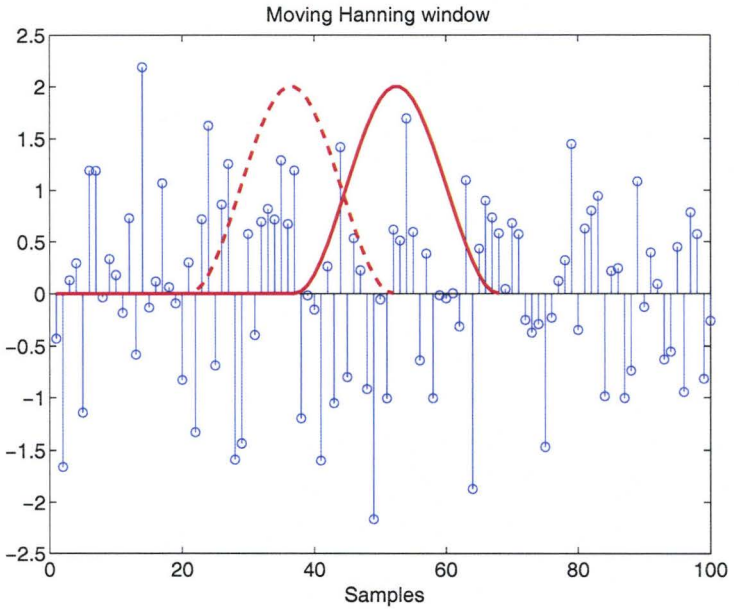

Figure 3.7: Example of a moving Hanning Window. The window is advanced 64 samples for subsequent frames. This provides 50 % data overlap in the frequency analysis of speech.

specified by *equivalent rectangular bandwidth* (ERB) scale [Glasberg & Moore, 1990]. The ERB is a function of the filter's center frequency fc, and is given by Eqn. 3.16.

$$ERB(f_c) = 0.108 f_c + 24.7 \tag{3.16}$$

Table 3.1 shows the center frequencies of each of the 15 bandpass filters used in the filterbank. The center frequency of each band is 1/3 octave apart starting from 250 Hz. The lowest band of the filterbank is a lowpass filter with cutoff frequency equal to 250 Hz. The bandwidth of each band is set to approximately 4 ERBs, i.e., 2 ERBs on either side of the center frequency. This allows a reasonable overlap of channels to minimize spectral ripple for tonal stimuli.

The amplitude response of each bandpass filter is determined by calculating the number of frequency bins in each filter. The size of frequency bin is 31.25 Hz (16 kHz/512 samples). The amplitude response in each channel is unity as shown in Figure 3.9 with a rolloff of 30 dB/octave. The extent of the rolloff is one octave above and below the unity-gain region. The spectral power P in each band for each frame is determined by,

$$P = (|H(\omega)|)^2.(|X(\omega)|)^2, \tag{3.17}$$

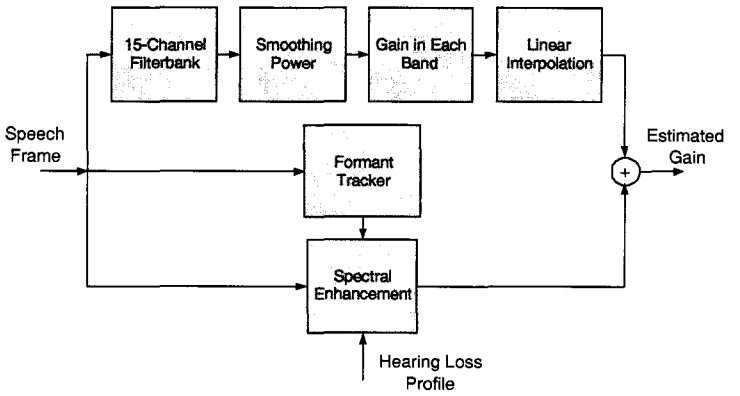

Figure 3.8: Schematic diagram of frequency domain gain (dB) calculation. The multiband compression and spectral enhancement are done in parallel. The total gain, that is the sum of the gains from multiband compressor and spectral-enhancement filter, is then applied to the original speech spectrum.

Table 3.6: Center frequencies of 15 bands in filerbank

Channel	Center Frequency (Hz)	Channel	Center Frequency (Hz)
1	250	9	1587
2	315	10	2000
3	397	11	2520
4	500	12	3175
5	630	13	4000
6	794	14	5040
7	1000	15	6350
8	1260		

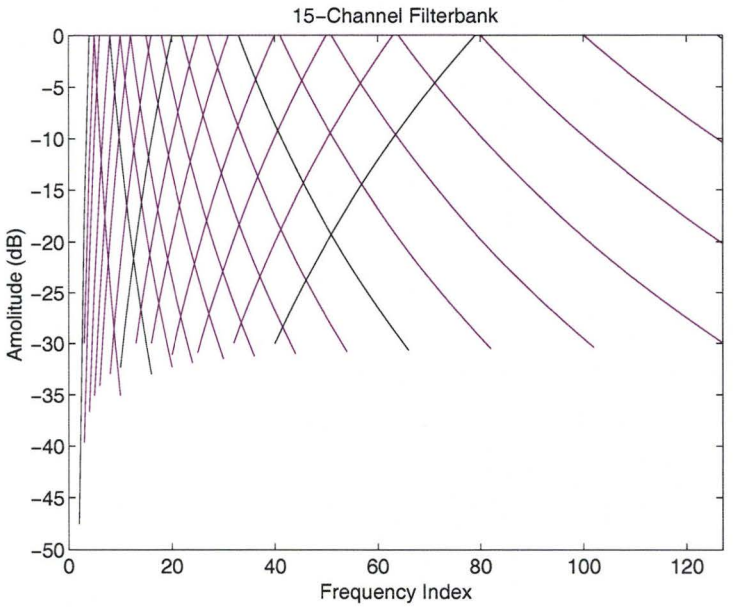

Figure 3.9: A 15-channel filterbank used in multiband compression in MICEFS. The center frequency of each bandpass filter is one-third octave apart. The passband of each bandpass filter is equal to 4 ERBs of the filter.

where $|H(\omega)|$ is the absolute value of the amplitude response of each filter in the filterbank, and $|X(\omega)|$ is the absolute value of the amplitude response of the input speech signal in each frame.

In MICEFS, *fast acting* or *syllabic compression* is used for hearing-impaired people with a very reduced dynamic range of hearing. In such compression schemes the time constants (attack and release times) are short. The release time in syllabic compression is less than the average length of a syllable in conventional speech, which is 200 to 300 ms (American English). In fast-acting compressors, the release time ranges from 50 to 150 ms. For very short release times, the compression system is referred to as *phonemic compression*, where intensive phonemes are compressed and weaker phonemes are amplified. In MICEFS, the attack time of compression is instantaneous, and the release time is approximately 60 ms. The release time is determined by using a single-pole IIR lowpass filter with pole at 0.68 and dc gain of 0.32. Figure 3.10 shows the impulse response of the IIR filter.

The *compression threshold* (or *kneepoint*) and *compression ratio* of fast acting compressors are assigned low values. The compression threshold is set at 40 dB SPL of the input level, which corresponds to softer sounds. The low compressor threshold ensures that the compression is on during most of the conversational speech. A fixed 2:1 compression ratio is applied in all frequency bands. In multiband compressors,

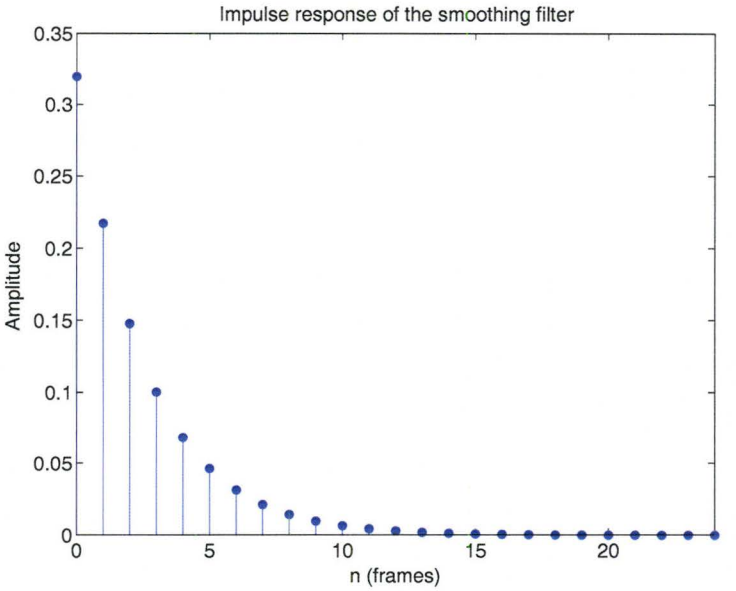

Figure 3.10: Single-pole IIR filter used in smoothing the power in multiband compression. The filter results in instantaneous attack time and a release time of 60 ms.

the static characteristics may vary in each band based on individual need. A gain of 0 dB is assigned for the input level below the kneepoint. Eqn. 3.18 is used for computing gain at the center frequency f_c of each band.

$$\text{Gain}(f_c) = G50 + (P - 50)\frac{G80 - G50}{80 - 50}, \tag{3.18}$$

where P is the input power in each band, and G80 and G50 are the gains at 50 and 80 dB SPL of the input level.

There are 512 frequency bins in each frame of the speech signal. The frequency bins from 257 to 512 are just the mirror image of the first 256 bins. The compression gain is evaluated at the center frequency of each bandpass filter in the filterbank (15 bins). The gain in other bins of the first 256 bins are determined by using linear *interpolation* and *extrapolation*. The gain at the second half of the frame is evaluated by mirror imaging the first half of that frame.

Spectral Enhancement using Time-Varying FIR Filter

A spectral enhancement filter is used in parallel with multiband compressor as

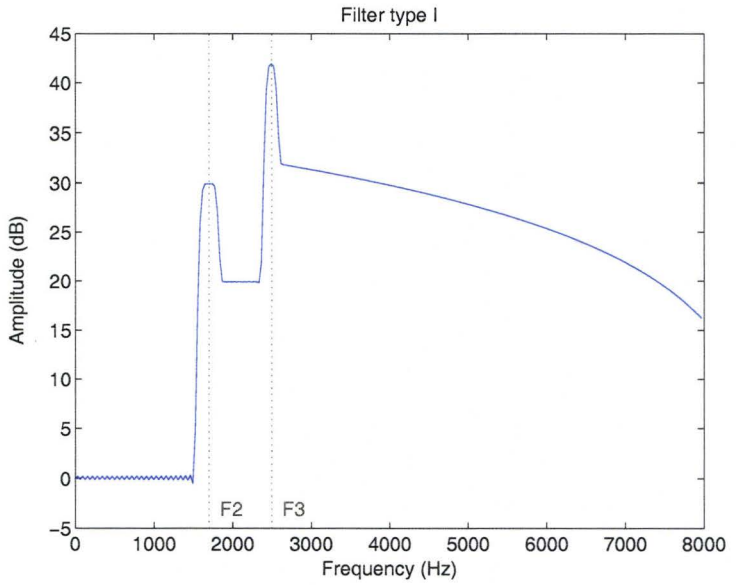

Figure 3.11: Time-varying FIR filter used in spectral enhancement. The frequency response of the filter is determined from formants of the voiced speech estimated by using a formant tracker and hearing-loss profile of impaired ear.

shown in Figure 3.8. The spectral enhancement is realized by using a time-varying FIR filter. The frequency response of the filter is determined using the average formants in each frame estimated by the formant tracker in real time and the audiogram of a hearing-impaired individual. A typical response of the spectral enhancement filter for a vowel signal is shown in Figure 3.11. The amplitude response of the filter has two narrowband peaks centered at F2 and F3. The peak at F3 is about 15 dB above the peak at F2. The peaks enhance the synchrony capture of the fibers with BFs in the proximity of F2 and F3. There is also some gain applied at the trough between F2 and F3 and at frequencies above F3 to improve the neural representation of the unvoiced components of speech. The gain at F2 is determined by the half-gain rule [Lybarger, 1944].

3.3.4 Overlap-And-Add Method

The combined gain thus obtained from the frequency domain linear filtering is then applied to the speech spectrum of each frame. In order to resynthesize the speech signal, the modified speech spectrum is transformed to the time domain by using the inverse Fourier transform (IFFT). Eqn. 3.19 shows a time domain sequence $\hat{x}[n]$ evaluated from IFFT of $\hat{X}[k]$.

$$\hat{x}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}[k]e^{j2\pi nk/N}, \qquad 0 \leq n \leq N-1 \tag{3.19}$$

The time domain frames of the speech signal are then added together using the overlap-and-add (OLA) method.

3.3.5 Data Representation

The model of the auditory periphery [Bruce et al., 2003] requires an input signal that is sampled at 500 kHz. The signal should be normalized to 0 dB SPL by the highest formant level before it is presented to the model. The speech sequence is repeatedly processed 100 times with a repetition rate twice the length of the stimulus. The neural data is recorded for 40 fibers with logarithmically spaced BFs from 0.1 to 5 kHz. The output is computed as the average poststimulus time histogram (PSTH) with bin size equal to 0.1 ms. A moving Hamming window $w(n)$ of 25.6 ms long is used to analyze the PSTH $p(n)$. $p(n)$ is then normalized to have units of spikes/sec. The synchronized rate is calculated by taking the Fourier transform of the windowed $p(n)$ as shown below [Bruce, 2004]:

$$R(m,k) = \left|\sum_{n=0}^{255} w(n)p(n+m)e^{-j2\pi nk/256}\right| \Bigg/ \sqrt{256\sum_{n=0}^{255} w(n)^2}, 0 \leq k \leq 255 \tag{3.20}$$

where m is the bin number of PSTH, k is frequency bin and n is the advancement of the moving Hamming window. The frequency resolution of STFT for a window length of 25.6 ms is $39\frac{1}{16}$ Hz. The denominator of the Eqn. 3.20 is the correction factor for the attenuation of the PSTH energy by the window function $w(n)$.

The quantification of the synchrony of the fibers to the formant frequencies F_x of the stimulus for a window starting at sample n is given as the power ratio (PR) of a formant. Spectral smearing caused by the windowing function is accounted for by including the synchronized rates from frequency components to each side of the exact formant frequency. If the frequency of F_x at sample index n exactly matches an STFT frequency component, i.e., is an integer multiple of $39\frac{1}{16}$ Hz, the PR includes the synchronized rates from the two frequency components of each side of the exact formant frequency giving a total of three frequency components with indices given by the vector $\mathbf{k}_x[n]$. If the frequency of F_x does not exactly match the STFT frequency components then the PR is calculated by including the synchronized rates from the two components of each side of the exact formant giving a total of four frequency components. The PR at sample index n for formant frequency F_x is given as [Bruce, 2004]:

$$\mathrm{PR}[n, F_x] = \frac{\Sigma_{\mathbf{k}_x} \mathrm{R}^2[n, \mathbf{k}_x]}{\Sigma_{k=2}^{255} \mathrm{R}^2[n, k]}, \tag{3.21}$$

The PR with a value equal to 0 corresponds to no synchrony and a value of 1 corresponds to a complete synchrony of the fibers to the formants. The formant frequency F_x at sample n takes into account that (a) the Hamming window is centered at sample $n + 127$, and (b) the average latency of model auditory nerve discharges is approximately 5 ms. The denominator of Eqn. 3.21 does not include the synchronized rate components for $k = 0$ or 1, which correspond to the average discharge rate of the model fiber.

Chapter 4

Simulation Results

MICEFS is a combination of multiband compression and improved contrast-enhancing frequency shaping intended to improve two important aspects of hearing loss: the reduced dynamic range of hearing and the diminished speech intelligibility respectively. Multiband compression has been used commercially in hearing aids to compensate for the reduced dynamic range of hearing. The CEFS algorithm is proposed to restore the normal neural representation of a vowel in an impaired cochlea [Bruce, 2004]. The focus of this thesis was to combine multiband compression with an improved version of CEFS. The improved CEFS was to found enhance the neural representation of voiced speech.

The MICEFS algorithm was tested using a computational model of the auditory periphery [Bruce et al., 2003] for two types of synthetic speech: a vowel and a sentence. The tests were conducted for forty fibers whose BFs were logarithmically distributed from 0.1 to 10 kHz. The average discharge rate in spikes/sec for each fiber was computed in response to the stimuli. The output of the model was plotted to analyze the average discharge rate of the fibers, and the synchrony capture of the fibers to the formants of the voiced speech. The synchrony capture phenomena in response to vowel was represented by box plots, which display solid squares of different sizes corresponding to the synchronized rate of the fibers. In the case of a sentence, power ratios of formants to those of the stimulus were plotted for the first three formants of the voiced speech. The power ratio plots quantified the synchrony capture of the fibers by the first three formants of the test sentence. The average short-term discharge rate of the fibers as a function of time using a 25.6 ms long Hamming window was represented by *neurogram* in response to the sentence.

In the MICEFS testing paradigm, the plots are compared for four different conditions:

1. unprocessed normal speech,

2. speech amplified according to the NAL-R prescription formula,

3. speech modified by CEFS,

4. speech modified by MICEFS.

In addition to the above conditions, MICEFS was also tested for formant estimates in each frame with some added noise with standard deviation of 100, 200 and 300 Hz. This test allowed for evaluation of the robustness of MICEFS in a noisy environment where there may be errors in formant estimation [Mustafa & Bruce, In Press].

4.1 Analysis of Unprocessed Speech to Normal Ear

Figure 4.1 shows the line spectrum of the vowel /ɛ/ used in the testing of MICEFS. The red curve shows the envelope of the spectrum with peaks at the formant frequencies. The vertical dashed lines are drawn to identify the first three formants of the vowel. The fundamental frequency or pitch of the vowel is 100 Hz. The first three formants are the 5th, 17th and 25th harmonics of the vowel respectively. The first formant dominates in spectral energy and the energy of the subsequent formants decreases gradually to zero.

The synchrony capture of the fibers in a normal ear in response to the vowel is characterized by the box plot as shown in Figure 4.2. The horizontal axis shows the best frequency of the fibers, and the vertical axis shows the frequency of the input vowel on a logarithmic scale. The diagonal gray bar describes the normal extent of the synchrony of the fibers to the formant frequencies, which is half an octave. Fibers with BFs equal to the formant frequencies are shown by vertical dashed lines. The formants of the vowel are shown by horizontal dashed lines. The level of the input vowel presented to the normal ear is 65 dB SPL. The solid dark block depicts the synchronized rate of a fiber in response to the vowel. The size of the block ranges from 15 to 120 spikes/sec with steps of 15 spikes/sec. The fibers show strong synchrony near F1, F2 and F3 of the vowel. The fibers in the trough between the formants show rate suppression, which creates a contrast between formant and nonformant harmonics of the vowel.

The average discharge rate as a function of the BFs of the fibers is shown in Figure 4.4. The horizontal axis gives the BFs of the fibers and the vertical axis gives the average discharge rate ranging from 300 to 600 spikes/sec. The vertical dashed lines show the formants of the vowel. Distinct peaks of the average discharge rate can be observed at the formant frequencies.

Figure 4.5 is a spectrogram of the test sentence and its PR for the first three formant frequencies. The test sentence is presented at a level of 75 dB SPL. The color bar defines the color gradients for the spectrogram magnitude and the formant

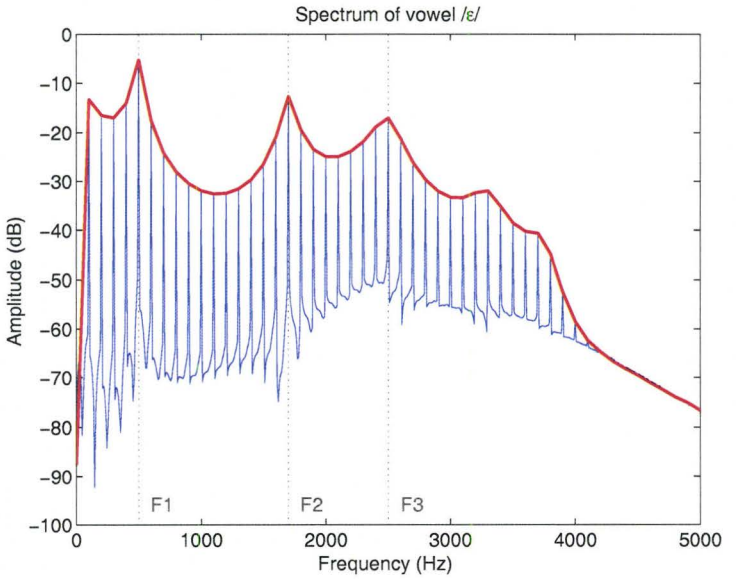

Figure 4.1: Line spectrum of the vowel /ɛ/ with envelope highlighted by the red curve. The vertical dashed lines show the first three formants of the vowel at 500, 1700 and 2500 Hz respectively.

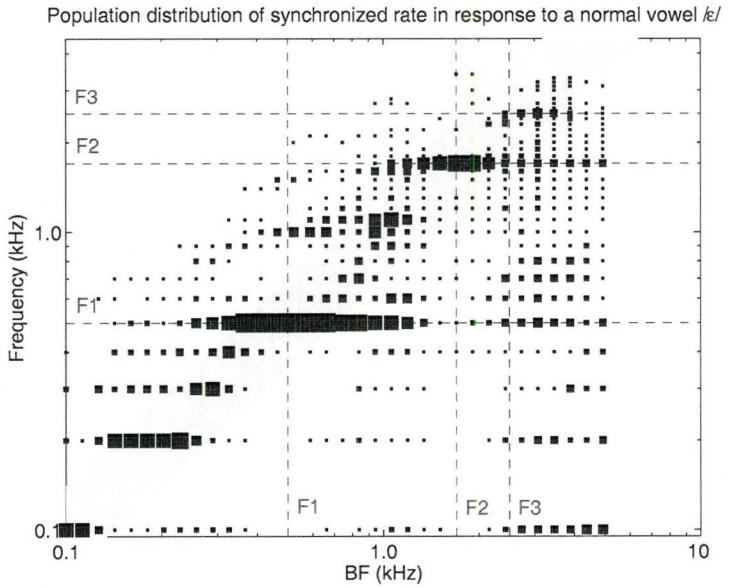

Figure 4.2: Box plot of the vowel /ɛ/ presented to a normal ear at 65 dB SPL. The fibers with BFs closer to the formants of the vowel demonstrate synchrony capture shown by the gray bar.

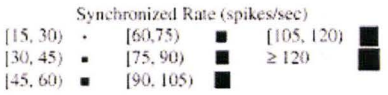

Figure 4.3: Legends for synchronized rate used in box plot in response to the vowel.

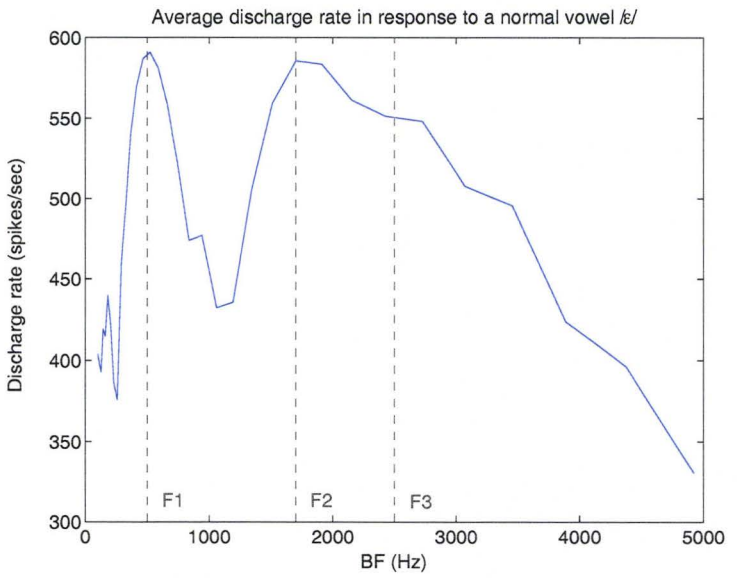

Figure 4.4: Average discharge rate of auditory nerve fibers in response to the vowel /ɛ/ presented to a normal ear at 65 dB SPL. The graph shows peaks at the formants of the vowel.

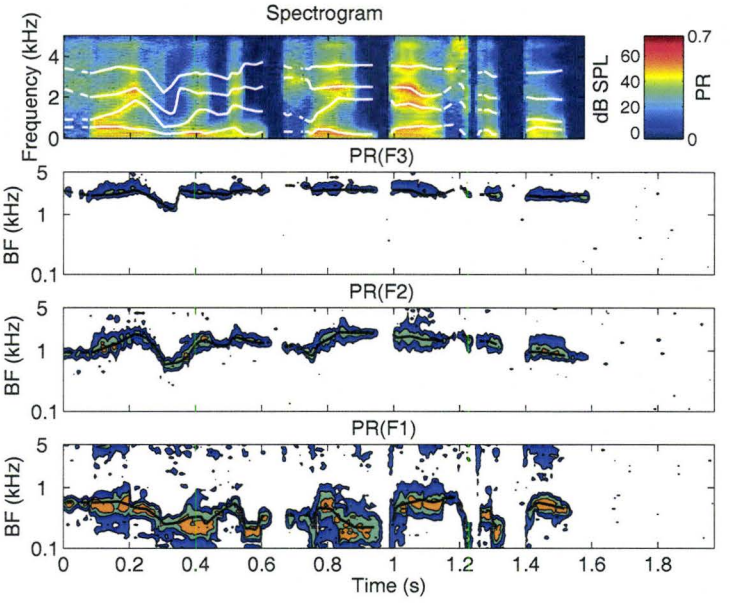

Figure 4.5: Spectrogram and PR plots of an unprocessed sentence presented to a normal ear at 75 dB SPL. The PR plots show the synchrony capture of the fibers to the formants of the sentence.

power ratios of the sentence. The horizontal axis in each subplot is the time axis and the vertical axis is the frequency on a linear scale or BF on a logarithmic scale. The solid lines in the spectrogram plot and in the power ratio plots show the formant trajectories of the voiced components of the test sentence. The dashed lines shown in the spectrogram and in the power ratio plots represent the absence of the formants for unvoiced speech. The discontinuity in lines represents the period of silence in the speech. The power ratio plots show a well defined synchrony of the fibers to formant frequencies of the voiced components.

The neurogram of the sentence is shown in Figure 4.6. The horizontal axis is the

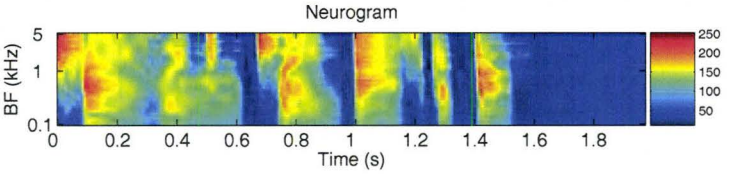

Figure 4.6: Neurogram of the unprocessed sentence presented to a normal ear at 75 dB SPL. The color bar shows the color gradients for the average discharge rate of the fibers.

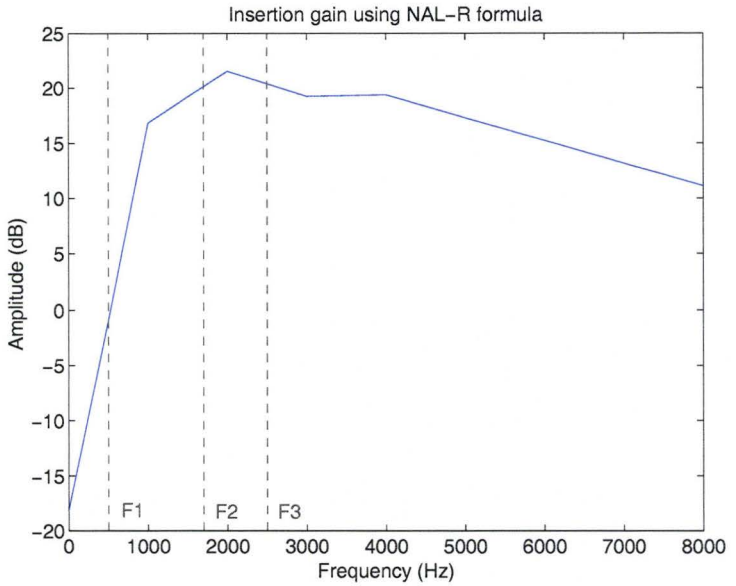

Figure 4.7: Gain-frequency response using NAL-R prescription formula. The low frequencies are attenuated and high frequencies are amplified to achieve loudness equalization in each of the frequency bands.

time axis, and the vertical axis is the BF on a logarithmic scale. The color bar defines the color gradients for the average discharge rate in response to the sentence measured in spikes/sec. Ideally, the neurogram should look like the spectrogram of the speech signal with a nonlinear frequency scale and the cochlear nonlinearities are taken into account.

4.2 Analysis of Modified Speech to Impaired Cochlea

4.2.1 NAL-R Prescription Formula

Figure 4.7 shows the insertion gain calculated using the NAL-R prescription formula. The speech signal is more intense at low frequencies. In order to achieve loudness equalization in all frequency bands, less gain is applied to low frequencies, and more gain is applied to high frequencies. Theoretically, loudness equalization should improve the intelligibility of speech for a hearing-impaired person. The vertical dashed lines show the formant frequencies of the vowel.
The box plot of the vowel modified by NAL-R and presented at 95 dB SPL is shown in Figure 4.8. The NAL-R amplification has restored synchrony capture of the fibers

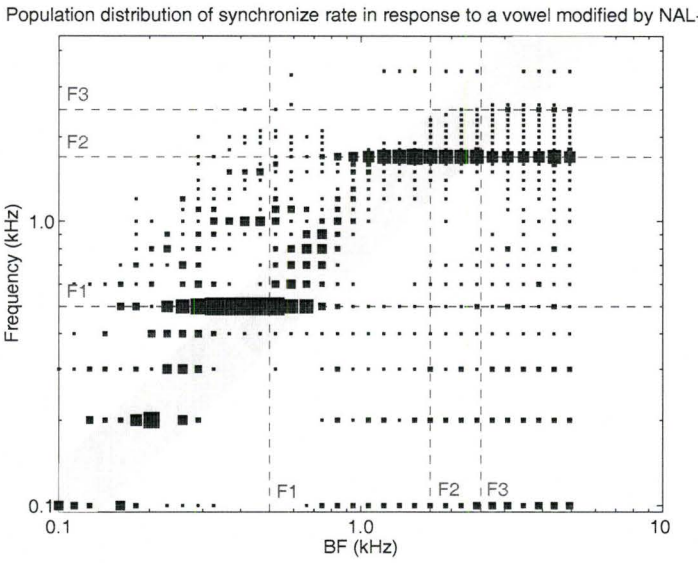

Figure 4.8: Box plot of vowel /ɛ/ amplified by NAL-R prescription formula and presented at 95 dB SPL. The fibers response to F1 is localized, but there is a broadband response to F2. The response to F3 is not restored.

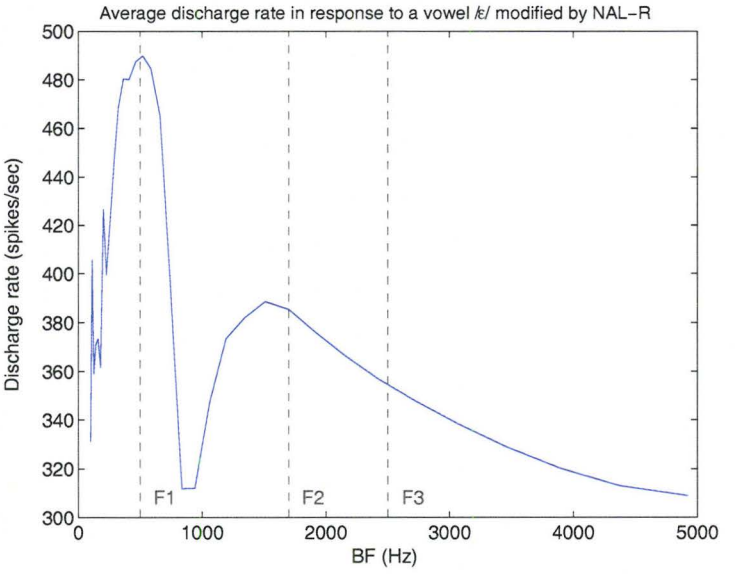

Figure 4.9: Average discharge rate of the fibers in response to vowel /ɛ/ amplified by NAL-R and presented at 95 dB SPL. The rate at F2 and F3 has not been restored, and there is no distinct peak at F3.

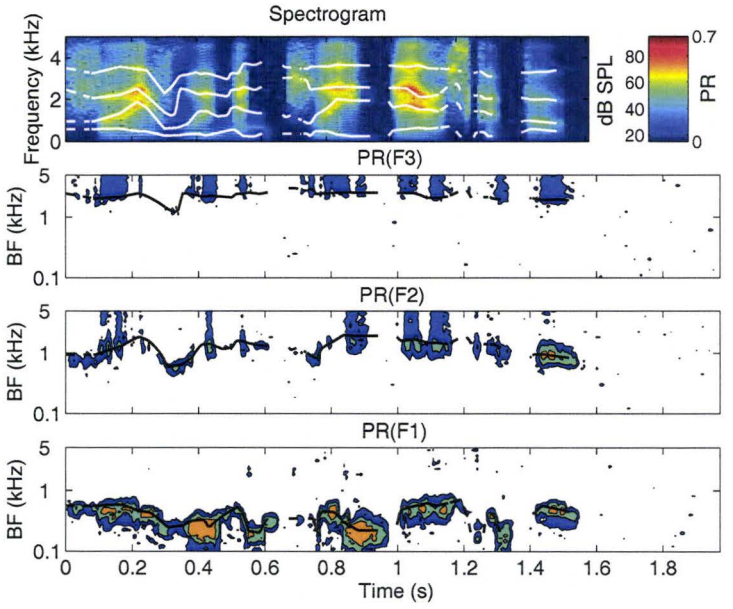

Figure 4.10: Spectrogram and power ratios of the formants to the total power of the stimulus amplified by NAL-R and presented at 95 dB SPL. The spectrogram shows some loss of spectral energy, especially below F1 and above F3. The synchrony of the fibers to F2 and F3 are very poor. There is also an upward spread of synchrony to F2.

with BFs near F1 and F2. The synchrony of the fibers at F1 is strong and well localized. The synchrony of the fibers at F2 is strong as well, however there appears to be an upward spread to the higher formants. Due to this upward spread, the synchrony of the fibers in the proximity of F3 has been lost. The average discharge rate of the fibers is plotted in Figure 4.9. There is a distinct peak at F1, but the rate has dropped about 100 spikes/sec relative to the average discharge rate of the fibers in a normal ear. The average discharge rate of the fibers near F2 has also dropped about 150 spikes/sec. Due to the upward spread of the response of the fibers with BFs close to F2, the average discharge rate of the fibers close to F3 is not distinct.

The spectrogram and the formant power ratio plots are shown in Figure 4.10. The magnitude response at frequencies below F1 has been diminished due to gain attenuation applied at low frequencies. The formant power ratio of F1 shows that the upward spread of the fibers' synchrony to F1 has been controlled. The localization of F1 synchrony has helped to restore synchrony of the fibers with BFs near F2. The power ratio of F3 is, however, very poor due to the upward spread of synchrony of the fibers with BFs in the proximity of F2. The neurogram in Figure 4.11 shows that there is loss of neural activity at high frequencies.

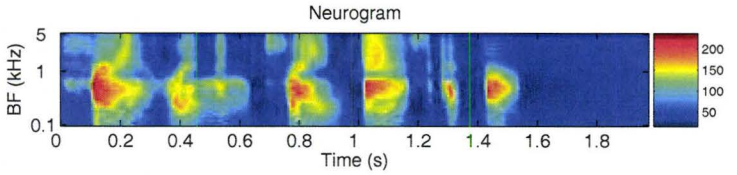

Figure 4.11: Neurogram of the speech sentence amplified by NAL-R and presented at 95 dB SPL. The plot shows some neural activity around F1. At high frequencies, the average discharge rate is lower as compared to the normal representation.

4.2.2 CEFS Modified Speech

The time-varying highpass filter used in CEFS is shown in Figure 4.12. The cutoff frequency of the filter is F2 less 50 Hz. The time-varying frequency response of the CEFS filter ensures amplification of harmonics of the speech signal near F2 without any amplification of the harmonics between F1 and F2. The passband gain is 30 dB, which is determined by the half-gain rule using the hearing-loss profile of the impaired ear. The gain drops smoothly after F3 to the square root of the gain at F2.

Figure 4.13 shows the population distribution of the synchronized rate in response to the CEFS-modified vowel /ɛ/ and presented at 95 dB SPL. The discharge rate pattern is almost the same as the response of the vowel processed by the NAL-R formula. In CEFS, however, the neural activity between F1 and F2 has been diminished because of the lack of amplification in the trough between F1 and F2. Figure 4.14 shows the average discharge rate of the fibers. The average discharge rate of the fibers with BFs near F2 shows a little improvement as compared to that of the NAL-R processed vowel. The increase in the average discharge rate at F2 could be due to rate suppression of the harmonics in the trough between F1 and F2. The average discharge rate of fibers near F3 is still indistinguishable due to the lack of fibers' synchrony close to F3.

The magnitude response of the non-formant harmonics below F2 is reduced as shown in Figure 4.15. At high frequencies, there is a diminishing magnitude response similar to that of NAL-R processed speech. The formant power ratio of F1 is large, and the synchrony of the fibers to F1 is localized. The fibers' synchrony to F2 is restored, but there is an upward spread of F2 synchrony to higher formants. Due to the upward spread of synchrony to F2, the fibers with BFs close to F3 do not exhibit any synchrony to F3. There is some loss of neural activity of the fibers at high frequencies as shown in Figure 4.16, which is consistent with the fibers' response to the NAL-R modified test sentence.

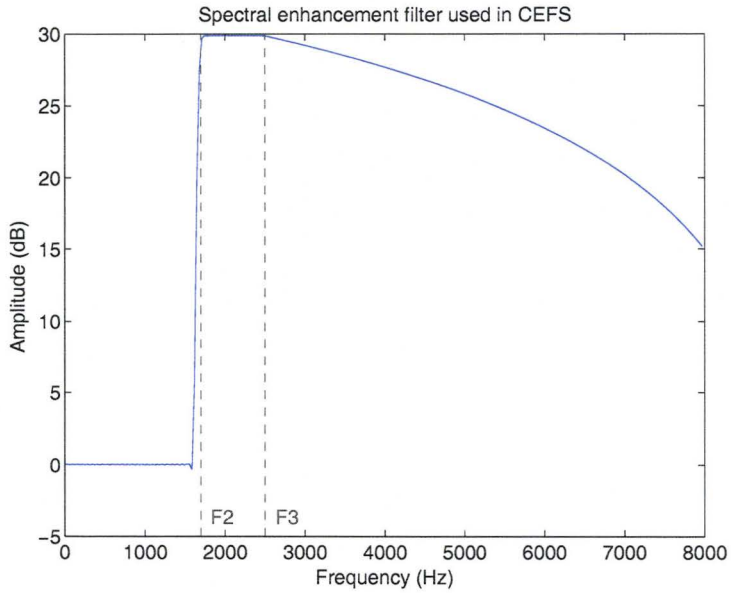

Figure 4.12: Frequency response of a time-varying FIR filter used in CEFS. The cutoff frequency is 50 Hz below F2 of a speech stimulus.

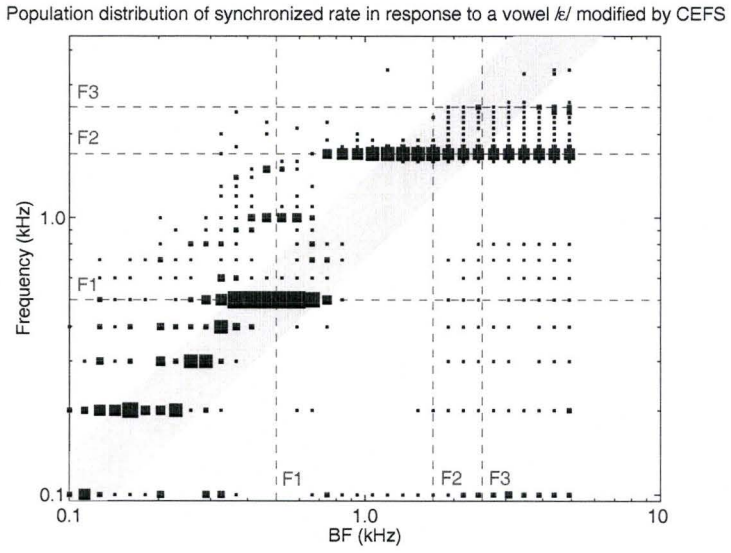

Figure 4.13: Box plot of CEFS-modified vowel /ɛ/ presented at 95 dB SPL. The fibers' synchrony to F1 and F2 have been restored. The synchrony capture of the fibers to F3 is still unresolved as expected from CEFS.

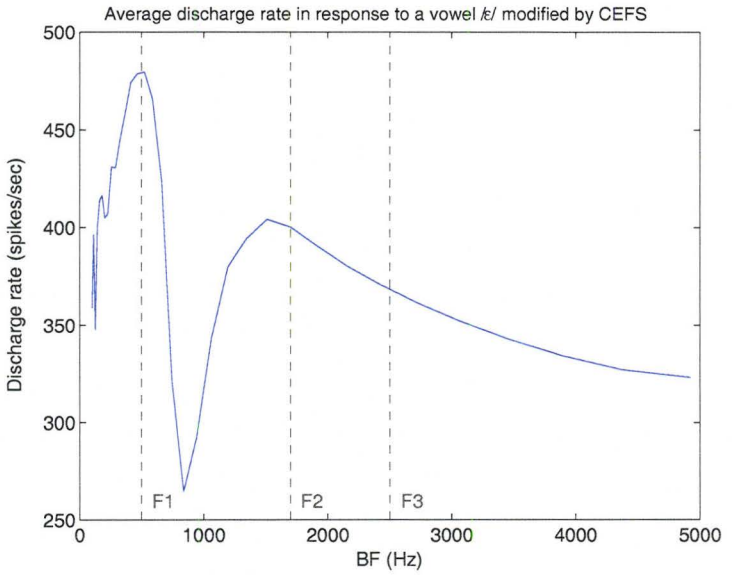

Figure 4.14: Average discharge rate of the fibers in response to CEFS-modified vowel /ɛ/ presented at 95 dB SPL. The neural activity at F2 and F3 is still low, and there is no obvious peak at F3.

4.2.3 MICEFS Modified Speech

The MICEFS algorithm is tested using two spectral-enhancement filters with different passband gains. The simulation results are given below for each filter.

MICEFS Filter I

The time-varying FIR filter of first type used for spectral enhancement is shown in Figure 4.17. In the case of a vowel, a 0 dB gain is applied from 0 to F2 less 150 Hz. Two narrow peaks with center frequencies at F2 and F3 are defined to have a bandwidth of approximately 200 and 100 Hz respectively. The height of the peaks are determined by the half-gain rule using the hearing-loss profile of an impaired ear. The frequency region between F2 and F3 is amplified with a gain 10 dB less than the gain at F2. The frequency region above F3 is amplified with a gain equal to 10 dB less than the gain at F3, which decreases smoothly to the square root of the gain at F3 minus 10 dB.

The box plot for MICEFS modified vowel /ɛ/ is shown in Figure 4.18. The input level of the speech stimulus to MICEFS was 75 dB SPL. The speech modified by MICEFS was presented to the model of a fiber at 95 dB SPL. The synchrony

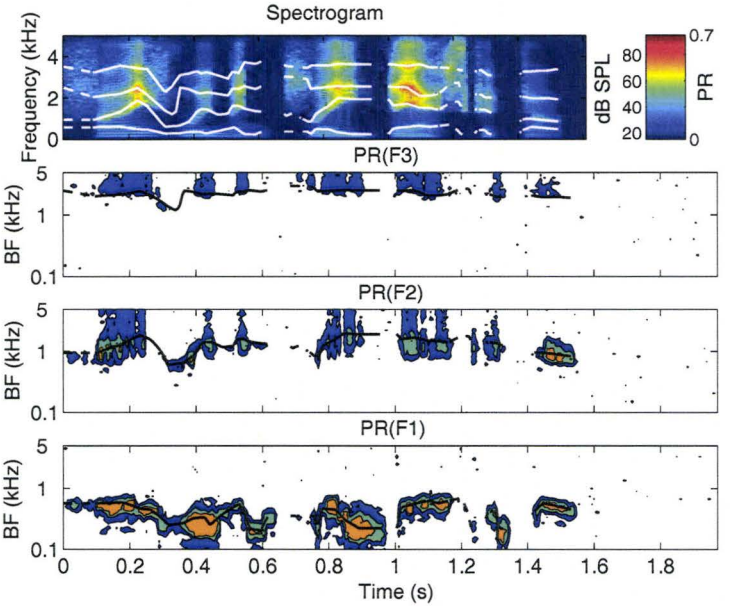

Figure 4.15: Spectrogram and formant power ratios of the sentence modified by CEFS and presented at 95 dB SPL. There is a loss in spectral energy below F2 and above F3 as compared to normal representation of the sentence. Synchrony to F1 and F2 have been restored. However, there is an upward spread of synchrony to F2.

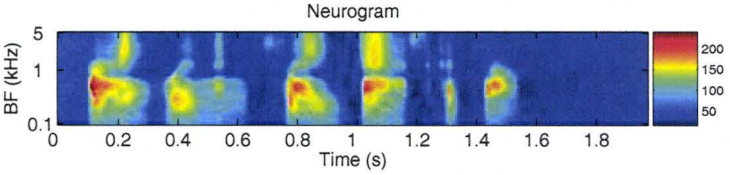

Figure 4.16: Neurogram of the sentence modified by CEFS and presented at 95 dB SPL. There is still a loss of neural activity at high frequencies above 1 kHz.

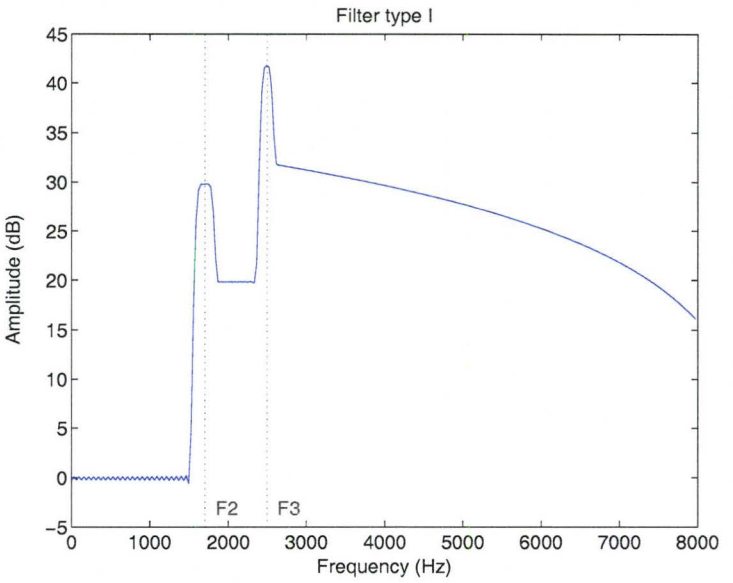

Figure 4.17: Time-varying FIR filter type I used in MICEFS. The narrowband peaks at F2 and F3 are intended to improve contrast between formant and nonformant harmonics of a speech stimulus.

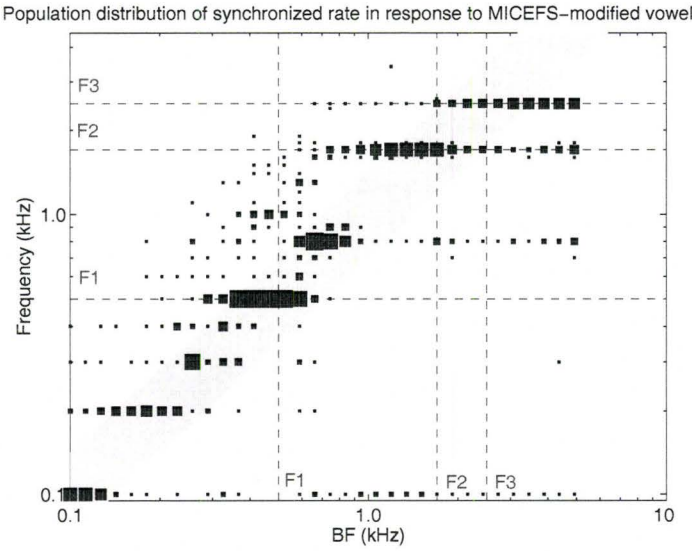

Figure 4.18: Box plot of MICEFS-modified vowel /ɛ/ using filter type I and presented at 95 dB SPL. The fibers' synchrony to F1, F2 and F3 have been restored. However, there is a distortion product at the 8th harmonic of the vowel.

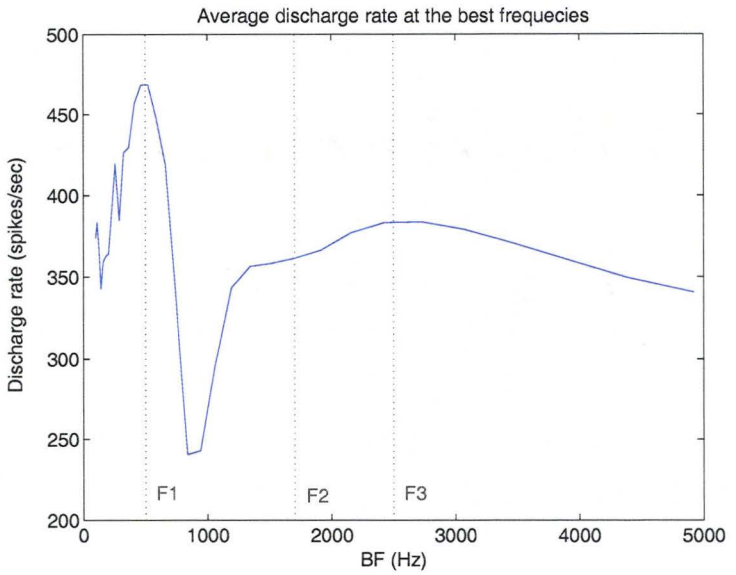

Figure 4.19: Average discharge rate of the fibers in response to MICEFS-modified vowel /ɛ/ using filter type I and presented at 95 dB SPL. The fibers still show loss of neural activity in response to the vowel at F2 and F3 as compared to normal representation of the vowel. However, there is an appearance of a peak at F3.

capture of the fibers near F1 is restored. The synchrony capture of the fibers to F2 has been localized to its proper place. There is no upward spread of synchrony of the fibers to F2, which results in restoring synchrony capture of the fibers in the proximity of F3. However, there is an upward spread of synchrony of the fibers to F3 due to high-frequency emphasis at F3. There is also a strong synchrony of the fibers to the eighth harmonic (800 Hz) of the vowel. This could be a distortion product from the difference of F3 (2500 Hz) and F2 (1700 Hz). The average discharge rate given in Figure 4.19 shows that there is slight loss of neural activity of the fibers to F1. The average discharge rate of the fibers near F2 is not prominent, and drops compared to CEFS. However, there is an improvement in the average discharge rate of the fibers close to F3 with some distinct peak.

The magnitude response in MICEFS has shown some improvement over CEFS in the lower frequency region around F1 as shown in Figure 4.20. The power ratio plots show a significant improvement in the synchrony to F1, F2 and F3 as compared to CEFS and NAL-R processed speech. Particularly, the upward spread of fibers' synchrony to F2 has been controlled, and this has resulted in a high formant power ratio at F3. The neurogram in Figure 4.21 shows a lot of high frequency activity which are missing in both NAL-R and CEFS.

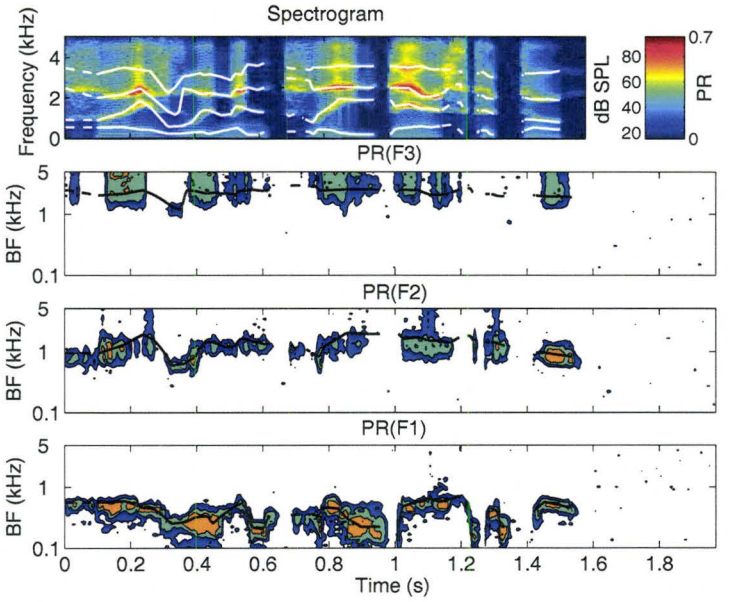

Figure 4.20: Spectrogram and formant power ratios of the sentence modified by MICEFS using filter type I and presented at 95 dB SPL. The spectrogram shows that the spectral energy at high frequencies has been restored. In the PR plots, the fibers' synchrony to F1, F2 and F3 have been improved.

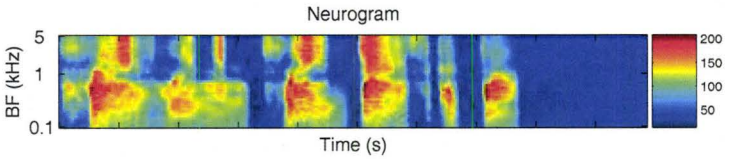

Figure 4.21: Neurogram of the stimulus modified by MICEFS using filter type I and presented at 95 dB SPL. The plot shows restoration of neural activity especially at high frequencies.

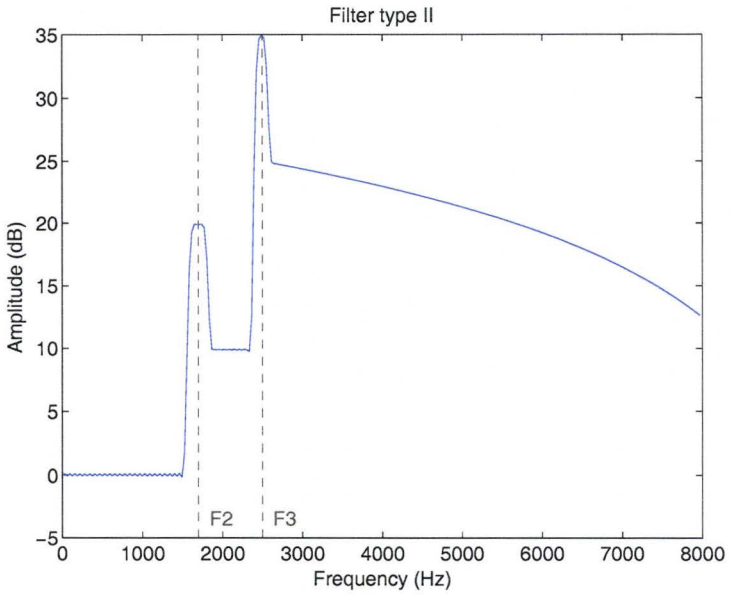

Figure 4.22: Time-varying FIR filter type II used in MICEFS for spectral enhancement. The passband gain is about 10 dB lower than filter type I.

MICEFS Filter II

The spectral enhancement filter of second type is shown in Figure 4.22. The only difference in filter type II is the passband gain, which is approximately 10 dB less than the filter type I.

Figure 4.23 shows the box plot for filter type II in response to a vowel sound. The plot shows that the distortion product at the eighth harmonic has subsided. Fibers of BFs close to F2 show a narrowband response, which has consequently improved the synchrony of the fibers near F3. The average discharge rate in Figure 4.24 also shows some significant improvement over filter type I of MICEFS. There is slight decrease in the average discharge rate of the fibers close to F1, however the rate at F2 and F3 has increased by about 100 spikes/sec. Another distinguishing feature in filter type II of MICEFS is that there are distinct peaks observed in the average discharge rate of the fibers near F2 and F3.

There is not much difference in the spectrogram shown in Figure 4.25 for filter type II over filter type I of MICEFS. The synchrony capture of the fibers near F2 and F3, however, shows a little improvement in filter type II. There is no obvious difference in the neurogram of filter type II as shown in Figure 4.26 when compared to that of filter type I of MICEFS.

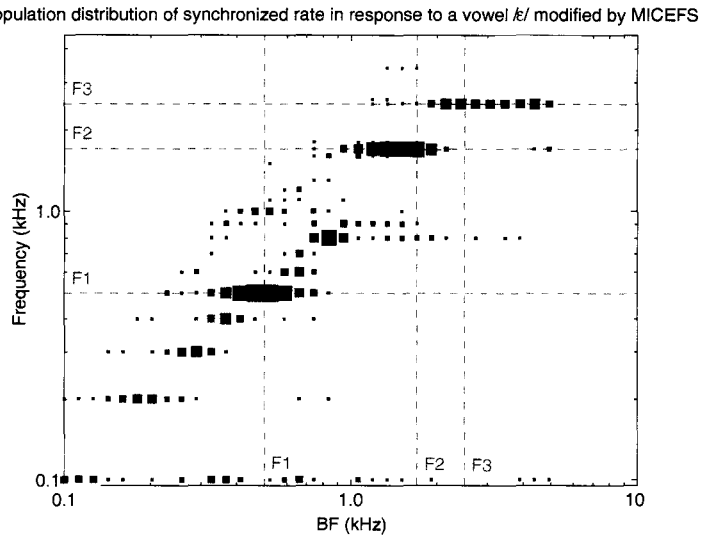

Figure 4.23: Box plot of MICEFS-modified vowel /ɛ/ using filter type II and presented at 95 dB SPL. The fibers show a little better localization of the synchrony to F2. Moreover, the distortion product at the eighth harmonic of the vowel is not very prominent.

4.3 MICEFS and Noisy Formant Estimates

The purpose of this test was to evaluate the performance of MICEFS in the presence of some error in the formant estimates for a speech signal. For moderate to low SNR, the error in formant estimation [Mustafa & Bruce, In Press] could misrepresent formants of the speech signal in each frame. In this test paradigm, an error was added to the formant estimates for each frame, where the error value was drawn from a zero-mean Gaussian distribution. Error distributions with standard deviations of 100, 200 and 300 Hz were evaluated.

4.3.1 Noise of 100 Hz Standard Deviation

Figure 4.27 shows the time-varying FIR filters with an error (standard deviation) of 100 Hz in the formant estimation for each frame. The vertical dashed lines show the actual formant frequencies for the vowel. The shift in formants affects bandwidth and the gain of the peaks at F2 and F3.
The neural responses of the vowel modified by MICEFS with noisy formants is represented in the box plot in Figure 4.28. There are no adverse effects on the synchrony capture of the fibers with BFs near the formants. It is interesting to note that the distortion product from the difference of F2 and F3 is also diminished. The

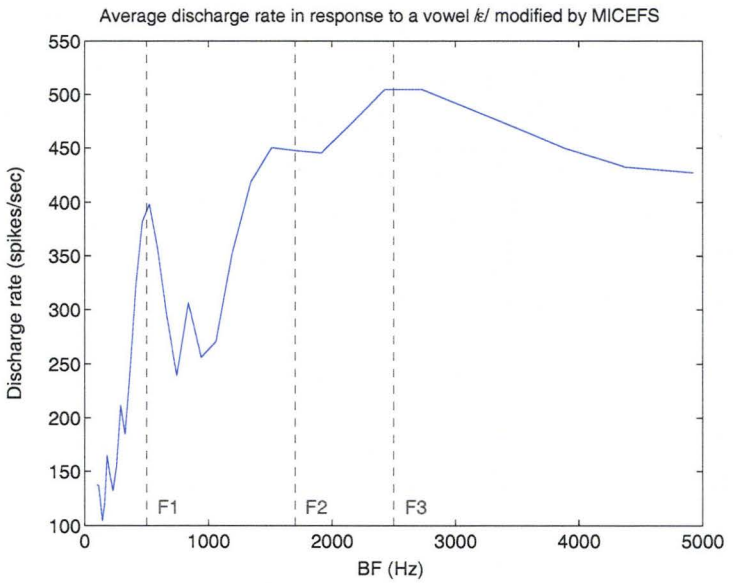

Figure 4.24: Average discharge rate of the fibers in response to MICEFS-modified vowel /ɛ/ using filter type II and presented at 95 dB SPL. There is a significant improvement of the neural activity of the fibers near F2 and F3 as compared to the one shown in filter type I. Also, there are prominent peaks of average discharge rate of the fibers close to F2 and F3.

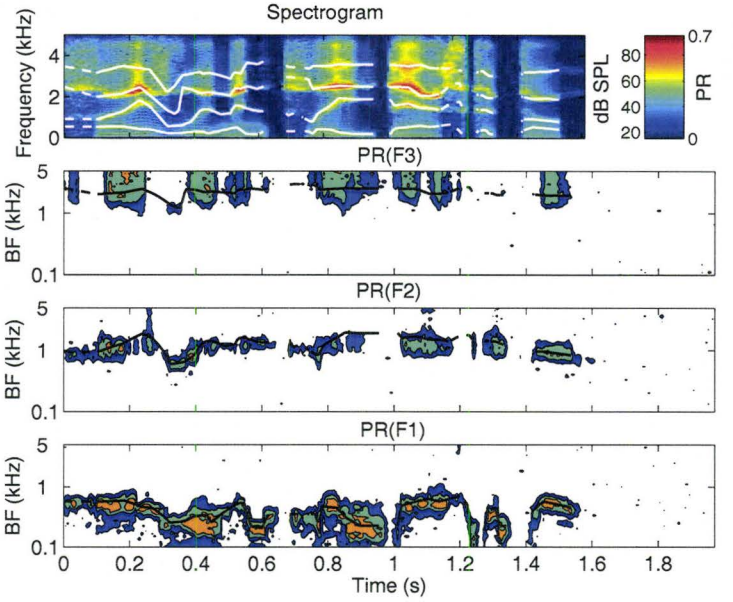

Figure 4.25: Spectrogram and formant power ratios of the sentence modified by MICEFS using filter type II and presented at 95 dB SPL. Spectrogram shows some improvement in spectral energy at high frequencies in relative to the one obtained in filter type I. The synchrony of the fibers to F2 has also improved.

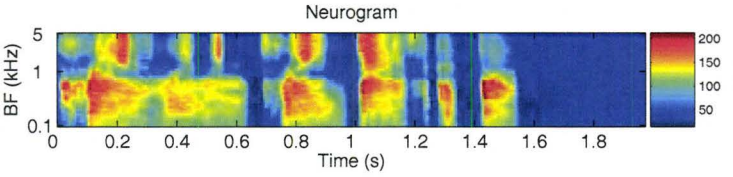

Figure 4.26: Neurogram of the fibers in response to the sentence modified by MICEFS using filter type II and presented at 95 dB SPL. The plot shows a similar response as was observed for filter type I.

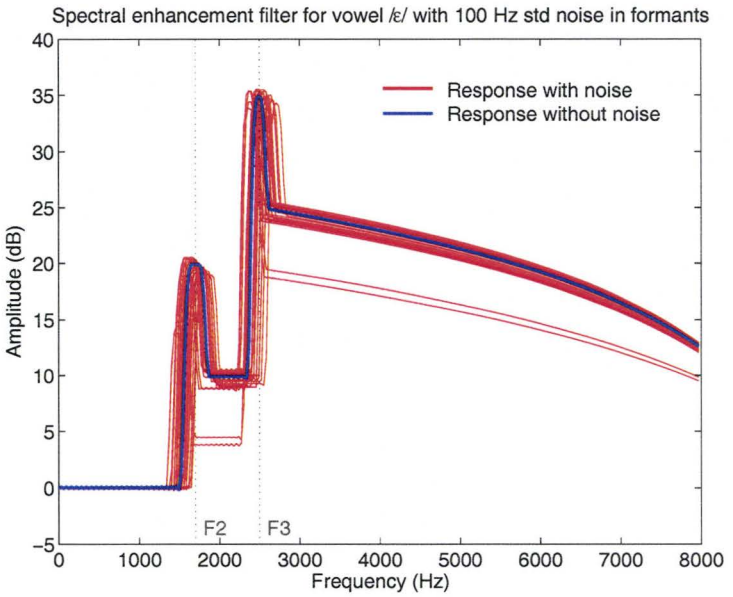

Figure 4.27: Example time-varying FIR-filter responses using formant estimates with added noise of 100 Hz.

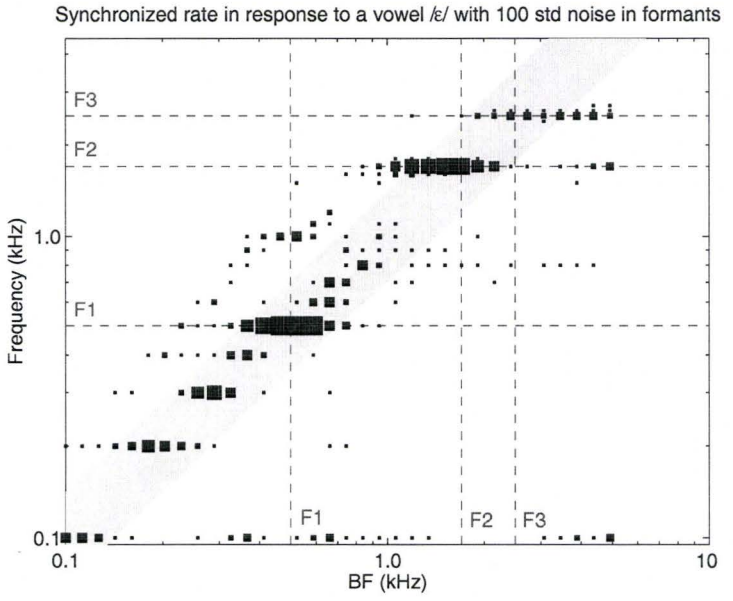

Figure 4.28: Box plot of MICEFS-modified vowel /ε/ with added noise of 100 Hz to the formant estimates for the vowel. The fibers' synchrony has not been affected.

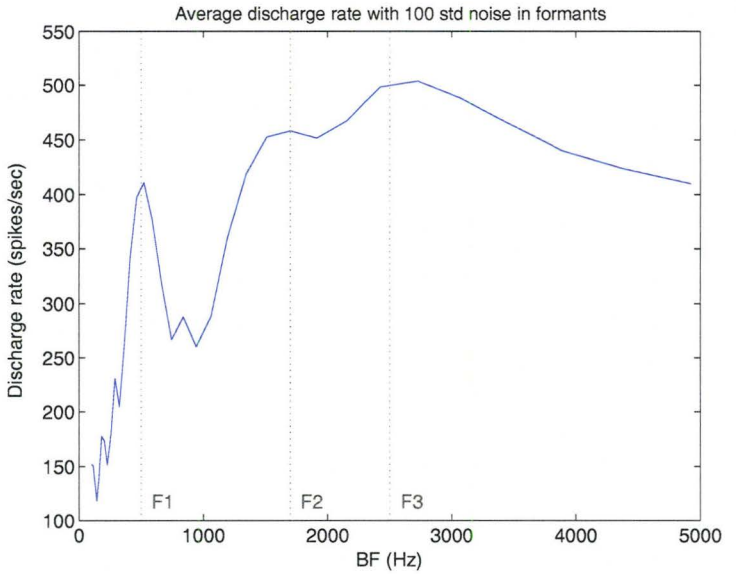

Figure 4.29: Average discharge rate of the fibers in response to MICEFS-modified vowel /ɛ/ with added noise of 100 Hz to the formant estimates. The plot shows that the neural activity of the fibers is unaffected.

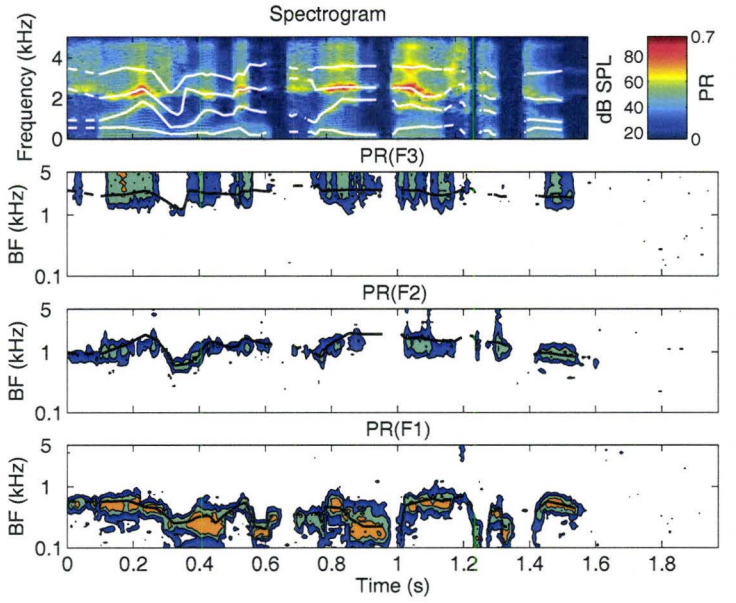

Figure 4.30: Spectrogram and formant power ratios of the stimulus modified by MICEFS with added noise of 100 Hz to the formant estimates. The spectral energy of the modified sentence and the fibers' synchrony are unaffected.

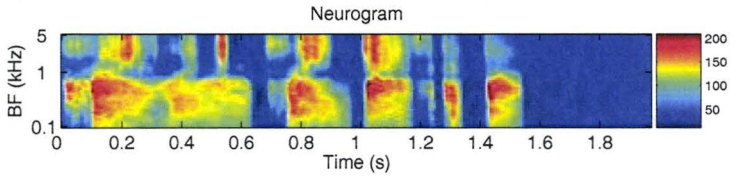

Figure 4.31: Neurogram of the stimulus modified by MICEFS with added noise of 100 Hz to the formant estimates. The neural activity of the fibers is not affected.

average discharge rate of the fibers in Figure 4.29 is also unaffected by noisy formant estimates.
The spectrogram and power ratio plots for the sentence in Figure 4.30 does not show any obvious difference in the presence of noise of 100 Hz in the formant estimates for the sentence. Similarly, the neurogram shown in Figure 4.31 describes the neural activity unvaried with noisy formant estimates for the speech signal.

4.3.2 Noise of 200 Hz Standard Deviation

The set of time-varying FIR filters with an error of 200 Hz is shown in Figure 4.32. The vertical dashed lines again show the actual formant frequencies of the vowel. The box plot in Figure 4.33 shows that the synchrony of the fibers with BFs near F3 is weakened significantly. The synchrony to F2 is also diminished. However, the synchrony to F1 is almost unaffected. The neural activity shown by the average discharge rate of the fibers in Figure 4.34 is still intact.
In the case of the sentence, the spectrogram and the power ratios in Figure 4.35 do not show any significant changes in the magnitude response. However, the synchrony to F2 and F3 has been diminished. The neural activity in response to the sentence has been reduced, especially at high frequencies as shown in Figure 4.36.

4.3.3 Noise of 300 Hz Standard Deviation

Figure 4.37 shows the set of amplitude responses of the time-varying FIR filters with an error of 300 standard noise added to the formant estimates.
The neural responses of the fibers with BFs in the proximity of F3 show some reduction in synchrony capture to F3 as shown in Figure 4.38. The fibers' synchrony to F2 has also been reduced. The fibers with BFs near F1, however, are still synchronized to F1 strongly. The average discharge rate of the fibers near F1 show a small increase in neural activity as can be seen in Figure 4.39. The average discharge rates of the fibers near F2 and F3 are not affected.
Figure 4.40 shows the spectrogram and the power ratio plots for the sentence. The magnitude response some deterioration, especially at the high frequencies. There is

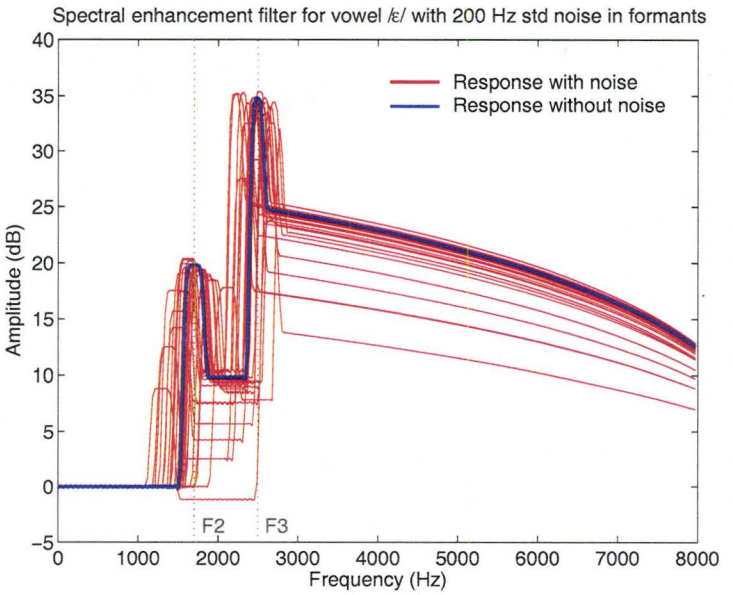

Figure 4.32: Example time-varying FIR-filter responses using the formant estimates with added noise of 200 Hz.

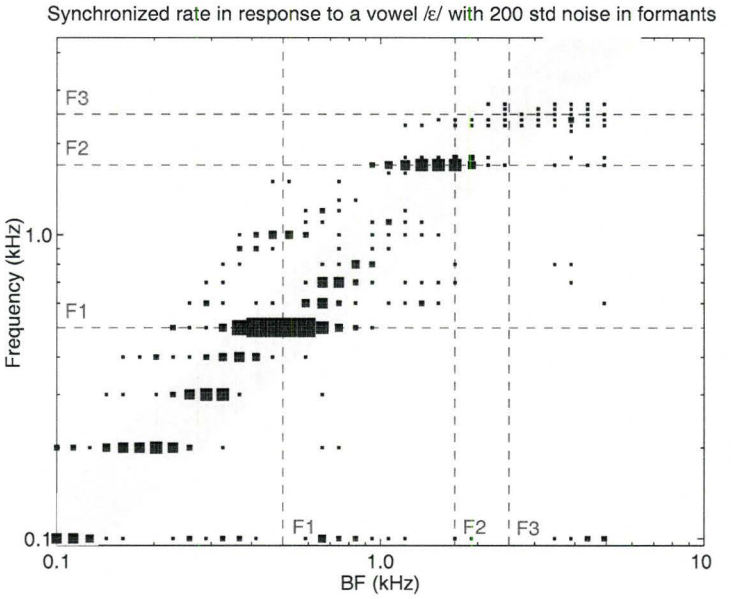

Figure 4.33: Box plot of MICEFS-modified vowel /ɛ/ with added noise of 200 Hz to the formant estimates for the vowel. The synchrony of the fibers to F3 has deteriorated.

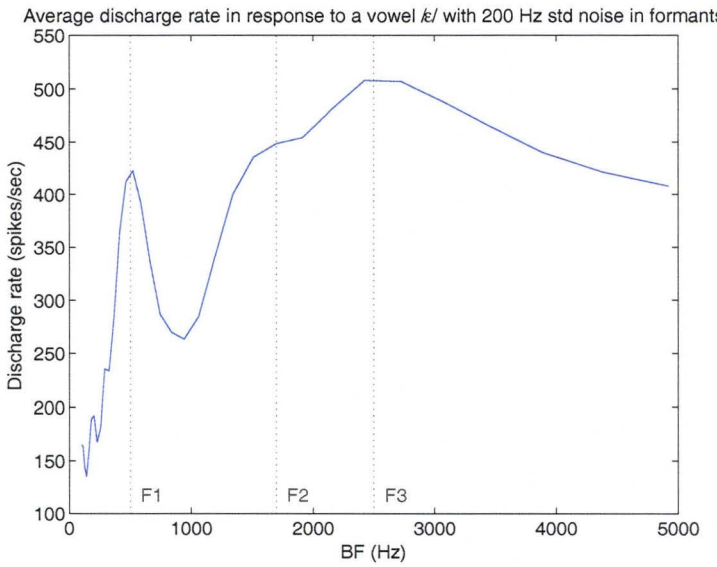

Figure 4.34: Average discharge rate of the fibers in response to MICEFS-modified vowel /ɛ/ with added noise of 200 Hz to the formant estimates for the vowel. The average discharge rate of the fibers has not been affected.

a noticeable loss of synchrony of the fibers to F2 and F3. The synchrony of the fibers to F1, however, is still strong. There is also a significant loss of neural activity at frequencies higher than 1 kHz as shown in the neurogram in Figure 4.41.

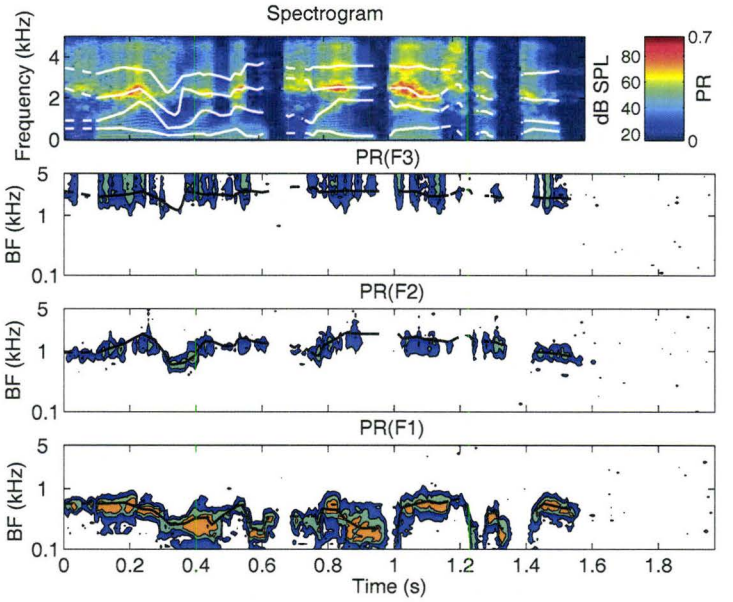

Figure 4.35: Spectrogram and formant power ratios of the stimulus modified by MICEFS with added noise of 200 Hz to the formant estimates for the sentence. The spectrogram is almost unaffected, however, the synchrony of the fibers to F2 and F3 lost as compared to the responses to MICEFS-modified sentence with no noise in formants.

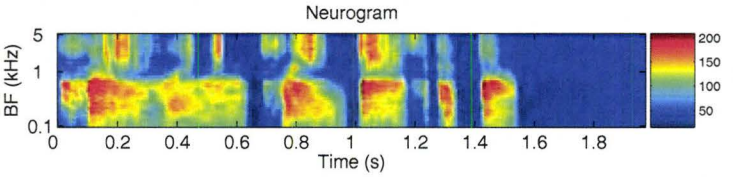

Figure 4.36: Neurogram of the stimulus modified by MICEFS with added noise of 200 Hz to the formant estimates for the sentence. There is some loss of neural activity around and above 1 kHz in relative to no noise response of MICEFS.

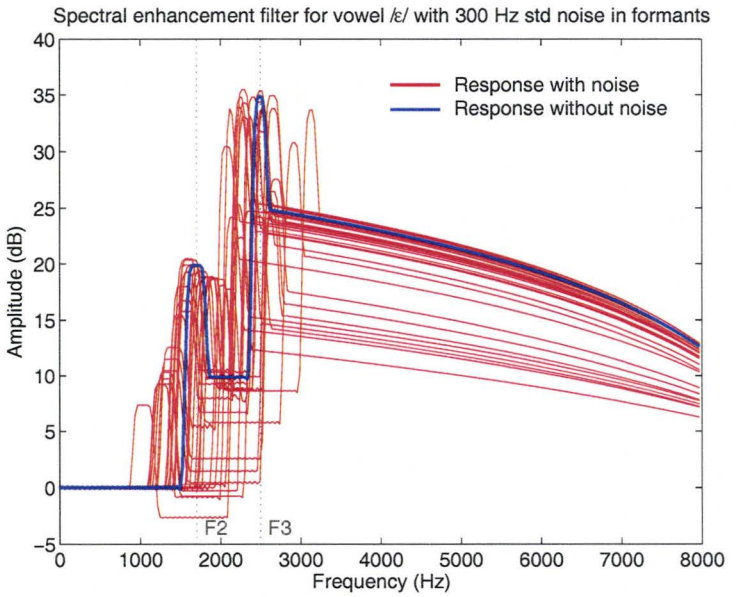

Figure 4.37: Time-varying FIR-filter responses using formant estimates with added noise of 300 Hz standard deviation.

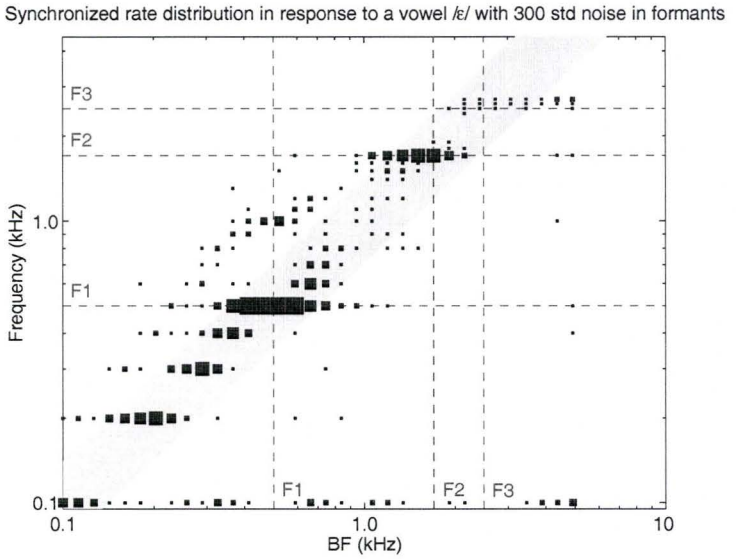

Figure 4.38: Box plot of MICEFS-modified vowel /ɛ/ with added noise of 300 Hz to the formant estimates for the vowel. The synchrony of the fibers close to F3 has been lost, whereas the synchrony to F2 has been diminished.

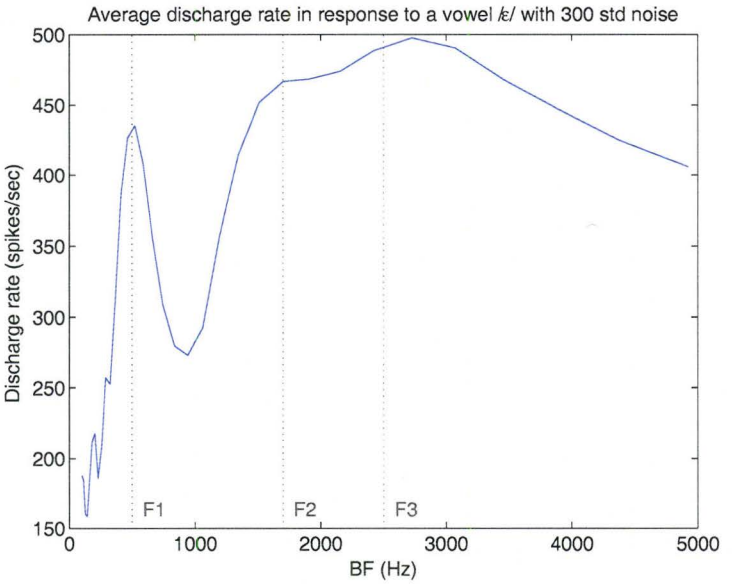

Figure 4.39: Average discharge rate of the fibers in response to MICEFS-modified vowel /ɛ/ with added noise of 300 Hz to the formant estimates for the vowel. The average discharge rate is not affected.

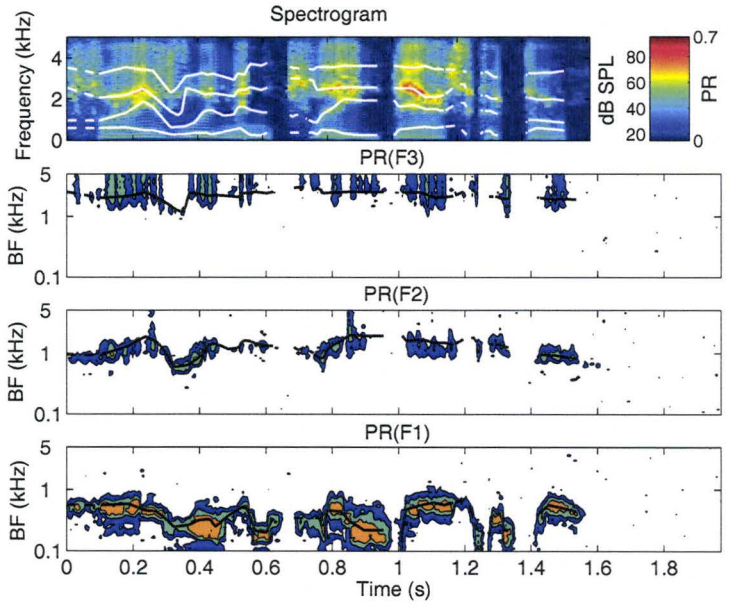

Figure 4.40: Spectrogram and formant power ratios of the stimulus modified by MICEFS with added noise of 300 Hz to the formant estimates for the stimulus. There is some loss of spectral energy above F2. The fibers' synchrony to F2 and F3 have been diminished.

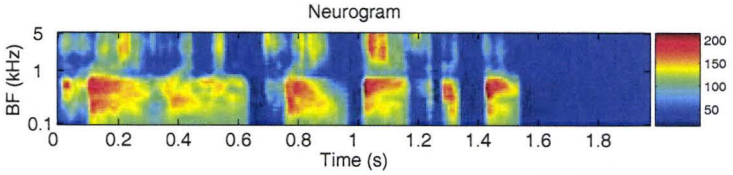

Figure 4.41: Neurogram of the stimulus modified by MICEFS with added noise of 300 Hz to the formant estimates for the sentence. The neural activity of the fibers above 1 kHz is diminished.

Chapter 5

Discussion

People with sensorineural hearing loss encounter loss of audibility and loss of frequency selectivity in speech communication. Physiologically, these losses can be related to the damage to OHCs and IHCs in the cochlea due to age, disease, drugs, or acoustic trauma. The auditory-nerve fibers show elevated and broadened tuning curves in an impaired cochlea. The psychophysical consequences of these effects are loss of audibility, constriction of the dynamic range and broadening of the auditory filters. The MICEFS algorithm is a hearing aid signal processing scheme, which has provided a combined solution for the reduced dynamic range of hearing and for reduced frequency resolution in hearing-impaired people. Multiband compression has been commercially used in hearing aids to match the dynamic range of speech with the reduced dynamic range of hearing of a hearing-impaired person. The method used in MICEFS for the compensation of reduced frequency selectivity in a hearing-impaired person is an improved version of CEFS, which is based on the neurophysiology of auditory nerve fibers.

5.1 Multiband Compression in MICEFS

In MICEFS, fast-acting multiband compression compensates for the reduced dynamic range of hearing. The reduction of the dynamic range of hearing can be explained by the loudness recruitment curve, where the rate of growth of loudness in hearing-impaired people is steeper than normal-hearing people as shown in Figure 5.1a. Figure 5.1b shows a typical compression scheme used to match the dynamic range of speech with the reduced dynamic range of hearing in a hearing-impaired person.

There are some disadvantages to using fast-acting multiband compression in hearing aids. The phonemic compression alters the intensity relationships between different phonemes and syllables. If a hearing-aid user uses the relative intensities of speech to help identify various phonemes, then altering relative intensities may deteriorate

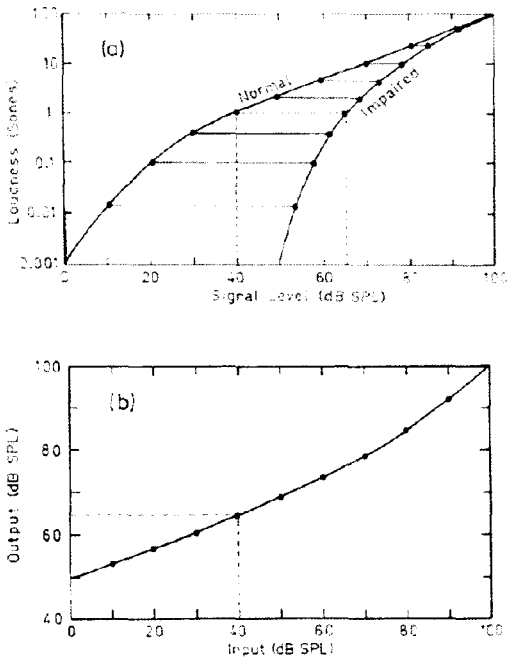

Figure 5.1: (a) Loudness growth curves for a person with normal hearing and with a sensorineural hearing loss, and (b) compressor input-output curve to provide normal loudness perception. Source: [Dillon, 1988].

intelligibility of some phonemes. The multiband compressors themselves are known to flatten the speech spectrum by giving less amplification to intense signals than to weak signals, which tends to decrease the height of spectral peaks and to raise the floor of spectral valleys. This spectral flattening makes it harder for the hearing-aid user to identify the place of articulation of consonants.

5.2 Formants Identification by Contrast Enhancement

The broadening of tuning curves of the auditory nerve fibers in an impaired cochlea has been related to damaged outer hair cells with some additional damage to inner hair cells. The broadening of tuning curves has direct implications for broadband fibers' synchrony to complex stimuli, such as speech. The analysis of phase-locking data shows that fibers with BFs near F2 and F3 in a damaged cochlea give responses to a wide band of stimulus including F1 and the harmonics between F1 and F2, and thus do not show synchrony capture to F2 or F3.

NAL-R

High-frequency emphasis used in the NAL-R formula [Byrne, 1986] can prevent the spread of F1 to higher BF fibers. This amplification scheme, however, also amplifies harmonics between F1 and F2 along with F2, which results in a broadband response of the fibers to F2. This shows that the NAL-R prescription formula partially resolves frequency selectivity in hearing-impaired people by improving synchrony capture of the fibers to F1 only. The sensitivity of the fibers to F2 and higher formants are still unresolved. The localization of F1 in NAL-R is achieved at the cost of spectral energy of the stimulus at frequencies below F2 where the gain of the stimulus is attenuated. The loss of sensitivity to F2 is also reflected on the neural activity of the fibers, which shows lower average discharge rate at F2 than the fibers in a normal ear.

CEFS

The CEFS algorithm [Miller et al., 1999] uses a time-varying highpass filter with a cutoff frequency just below F2 that provides gain to F2 and higher frequencies of stimulus without amplifying harmonics between F1 and F2. The passband gain is determined by the half-gain rule [Lybarger, 1944]. CEFS amplification improves the synchrony of the fibers to F2 along with F1. There is, however, upward spread of synchrony to F2, which masked the responses of the fibers to F3. The spectrogram of the CEFS-modified stimulus shows some loss of the spectral energy below F2, which can be expected by CEFS filtering. Although, CEFS is able to restore phase locking to F2, the average discharge rate of the fibers at F2 is still not restored.

MICEFS

The MICEFS algorithm is a combination of multiband compression and an improved version of contrast-enhancing frequency shaping. The improved version of CEFS restores the synchrony of the fibers in an impaired cochlea to F1, F2 and F3 of the stimulus. The synchrony to F1 and F2 are achieved by using a highpass filter similar to the one used in original CEFS. The only differences in the frequency responses of MICEFS and original CEFS are the two narrowband peaks in the passband region centered at F2 and F3. These narrowband peaks, in addition to improving the synchrony of the fibers to F2 and F3, restore the average discharge rate at F2 and F3. In the simulation of MICEFS using filter-type I, the peak at F2 is half of the hearing loss (30 dB) and the peak at F3 is about 12 dB above the peak at F2 (42 dB). In this simulation, the synchrony of the fibers at F1, F2 and F3 have been improved, and there is an improvement in the average discharge rate of the fibers at F2 and F3. However, there is also a distortion product in the fibers' responses, where the fibers seem to synchronize to the 8th harmonics of the stimulus. This eighth harmonic is the frequency difference of F2 and F3, and may have resulted from the amplification applied at F2 and F3.

In the simulation of MICEFS using filter-type II, the peak at F2 is lowered about 10 dB. This gives a gain of about one-third of the hearing loss at F2 (20 dB). The peak at F3 is lowered about 7 dB to a level of 35 dB that gives a level difference of 15 db between the peaks at F2 and F3. The results of filter-type II show an significant improvement, particularly, in the average discharge rate of the fibers at F2 and F3, which is achieved by increasing the relative gain at F2 and F3. Moreover, the results from MICEFS using filter-type II show that the distortion product at the 8th harmonic of the stimulus is no longer present in the neural responses of the fibers to the speech stimulus.

5.3 Robustness of MICEFS

MICEFS uses a time-varying FIR filter with arbitrary gain for spectral enhancement, which requires formant information of the speech stimulus and hearing loss profile of hearing-impaired person to determine its real-time frequency response. The accuracy of the formant information is very crucial for the performance of MICEFS, especially in running speech. A discrepancy in the estimated formants may distort the temporal representation of spectral shape of the stimulus at the auditory nerve fibers in hearing-impaired listeners. The MICEFS algorithm uses a robust formant tracker [Kamran & Bruce, In Press] for the real-time estimation of the formants of the speech stimulus. The formant tracker has been extensively tested for various speech environments, and proved to be the best choice for the purposes of implementing MICEFS. In order to evaluate the robustness of MICEFS a set of three tests were conducted with some Gaussian noise added deliberately to the estimated formants of the speech. The standard deviation of the noise in the three tests were set at 100, 200 and 300 Hz. The test results from noisy formants show that the synchrony to F2 and F3 deteriorated gradually with increase in noise in response to speech sounds. The results show that the noise has more adverse effects on synchrony to F3 than F1 and F2. The synchrony of the fibers to F1, however, seems to be least affected by the noise in formants. The discharge rate of the fibers, however, remains almost constant in all the three cases of noisy formants. This raises a research-worthy question: Is synchrony capture of the fibers with BFs close to the formants more important for the perception of speech, or the average discharge rate of the fibers?

5.4 Time Delay Response of MICEFS

Bruce [2004] implemented multiband compression in series with time-domain spectral enhancement filter. The total time delay resulted from the series implementation was about 26 ms. The MICEFS algorithm is implemented in frequency domain by using a 15-channel filterbank for multiband compression in parallel with a time-varying FIR filter for spectral enhancement. In MICEFS, the total time delay of the algorithm is reduced to 16 ms, which is a 10 ms improvement over the series implementation. MICEFS uses block or frame processing by windowing the input speech signal. The frame size in MICEFS is 512 samples, which produces a delay of 256 samples (half of the frame size) at the output. The impulse response of MICEFS has also been determined for the shorter frame size of 256 samples, but the result shows time aliasing in the response.
The time delay in MICEFS is still longer than the required standard for hearing aids, which is 10 ms. A time delay longer than 10 ms affects a hearing-aid user's

perception because of the combined effects of the amplified and non-amplified sound [Stone & Moore, 1999]. A delay of more than 40 ms may put the auditory information out of synchronization with visual information from lip-reading [McGrath & Summerfield, 1985; Summerfield, 1992]. Another problem with time delay is that amplified and non-amplified sounds may cancel at some particular frequencies and the likelihood of cancellation is greater for longer delays.

5.5 Implications for Hearing Aids

The internal spectrum of a vowel in hearing impaired subjects has been estimated in perceptual studies [Bacon & Brandt, 1982; Sidwell & Summerfield, 1985; Tasell et al., 1987]. The studies show a flattening of the internal spectrum as compared to the pattern observed in normal listeners. This flattening of the internal spectrum is generally assumed to arise from the broadening of auditory filters due to loss of OHCs and IHCs. The goal in MICEFS and in previous algorithms [Miller et al., 1999; Schilling et al., 1998] is not only to restore the audibility at high frequencies, but also to restore the neural representation of the speech stimulus. Other perceptual studies show that restoring audibility does not restore the speech perception performance to the extent predicted by the articulation index [Ching et al., 1998].

Previous studies have shown that at high input level, the internal representation of speech in hearing-impaired people is degenerated. The neural studies at high sound levels support this hypothesis [Wong et al., 1998]. This can also be confirmed from the responses of two spectral enhancement filters in MICEFS with slightly different passband gains. The results show that the average discharge rate has diminished as the passband gain is increased by about 10 dB. The perceptual literature shows that there is an upward spread of masking in hearing-impaired listeners [Danaher & Pickett, 1975; Summers & Leek, 1997]. The psychophysical interpretation of the increased upward masking is the broadened auditory filters, especially on the low frequency side [Tyler et al., 1984; Stelmachowicz et al., 1985; Glasberg & Moore, 1986]. This has a correlate with the broadening of the fibers' tuning curves also in the low frequency side [Dallos & Harris, 1978; Schmiedt et al., 1980; Liberman & Dodds, 984b]. This results in upward spreading of response to F1, which masks the higher formants. The upward spreading of F1 and masking of the higher formants can be controlled by high frequency emphasis.

Chapter 6

Conclusions

6.1 Concluding Remarks

In this thesis, a hearing-aid signal processing scheme is presented for people with reduced dynamic range of hearing and reduced frequency selectivity. The scheme is called MICEFS and is a combination of multiband compression and an improved CEFS. The multiband compression is realized by a 15-channel filterbank, which works in parallel with a spectral enhancement filter implemented by a time-varying FIR filter. A fast-acting compression, intended for people with very reduced dynamic range of hearing, is employed in each of the 15 bands. The multiband compression, however, tends to flatten the spectral shape of the speech stimulus, which may cause a difficulty for hearing-aid users to identify the place of articulation of consonants [Gennaro et al., 1986; Lindholm et al., 1988]. For smaller compression ratios, studies have shown that the multiband compression improves the speech audibility, and the effects spectral flattening are smaller [Yund & Buckles, 1995].
The spectral enhancement in MICEFS is achieved by a time-varying FIR filter with arbitrary gain. The frequency response of the filter is determined by using formants of the stimulus estimated in real-time by a formant tracker [Mustafa & Bruce, ress] and the audiogram of a hearing-impaired person. Improved CEFS stops the upward spread of fibers' synchrony to F2 and improves the phase locking of the fibers to F3. It has also restored the average discharge rate of the fibers at F2 and F3 in response to speech stimulus in an impaired cochlea. Improved CEFS, however, does not prevent the upward spread of fibers' synchrony to F3, which may mask the neural responses to higher formants. The frequency response of the time-varying FIR filter depends on the accuracy of the formants estimation by the formant tracker. Experiments employing MICEFS on noisy formant estimates have shown that estimates with an error of 200 Hz and more may have a devastating effects on the synchrony of the fibers particularly at F2 and F3.
The frequency domain implementation of MICEFS has improved the time delay

response of the algorithm as compared to the series implementation of multiband compression and time domain spectral enhancement filter [Bruce, 2004]. The time delay of MICEFS is about 16 ms, which is 10 ms less than its series implementation counterpart (26 ms). The time delay in MICEFS is still higher than standard time delay in hearing aids, which is 10 ms. However, if the MICEFS algorithm greatly improves the intelligibility of speech in hearing-impaired people, then this time delay of 16 ms may be tolerated.
The MICEFS algorithm has been tested using a model of the auditory periphery. The responses of the model to MICEFS-modified speech are promising in terms of synchrony capture and the average discharge rate of the auditory fibers. However, the actual performance of the MICEFS algorithm requires testing on human subjects.

6.2 Future Work

In future updates of MICEFS, the formant tracker may be incorporated into MICEFS, which may avoid any mismatch of the formant estimation whatsoever. The total time delay of the algorithm may be cut down further to meet the standard time delay in hearing aids (10 ms). The MICEFS algorithm improves contrast between formant and nonformant harmonics of a voiced speech signal. In future version of MICEFS, unvoiced speech may also be enhanced to improve the intelligibility of unvoiced consonants in hearing-impaired people. This may be done by using a real-time switch for the selecting a either spectral enhancement filter or linear gain filter for voiced and unvoiced speech respectively.

Bibliography

ANSI S3.22 (See Dillon, 2001).

Allen, J., Hall, J., & Jeng, P. (1990). Loudness growth in 1/2-octave bands (LGOB) - a procedure for the assessment of loudness. *J. Acoust. Soc. Am.*, *88*(2), 745–753.

Allen, J. B. (1977). Short term spectral analysis, synthesis and modifications by discrete fourier transform. *IEEE Trans. Acoust. Speech Sig. Proc.*, *25*, 235–238.

Bacon, S. & Brandt, J. F. (1982). Auditory processing of vowels by normal-hearing and hearing-impaired listensers. *J. Speech Hear. Res.*, *25*, 339–347.

Baer, T. & Moore, B. C. J. (1993). Effects of spectral smearing on the intelligibility of sentence in the presence of noise. *J. Acoust. Soc. Am.*, *94*, 1229–1241.

Boers, P. M. (1980). Formant enhancement of speech for listeners with sensorineural hearing loss. *IPO Ann. Prog. Rep.*, *15*, 21–28.

Boothroyd, A., Springer, S., Smith, L., & Schulman, J. (1988). Amplitude compression and profound hearing loss. *Journal of Speech and Hearing Research*, *31*, 362–376.

Braida, L. D., Durlach, N. I., Lippmann, R. P., Hicks, B. L., Rabinowitz, W. M., & Reed, C. M. (1979). Hearing aids: A review of past research on linear amplification, amplitude compression, and frequency lowering. *American Speech and Hearing Monograph*, *19*.

Bruce, I. C. (2004). Physiological assessment of contrast-enhancing frequency shaping and multiband compression in hearing aids. *Physiol. Meas.*, *25*, 945–956.

Bruce, I. C., Sachs, M. B., & Young, E. D. (2003). An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses. *J. Acoust. Soc. Am.*, *113*(1), 369–388.

Bunnell, H. T. (1990). On enhancement of spectral contrast in speech for hearing-impaired listeners. *J. Acoust. Soc. Am.*, *88*, 2546–2556.

Burns, W. (1973). *Noise and Man* (2nd ed.). London: John Murray.

Bustamante, D. K. & Braida, L. D. (1987). Multiband compression limiting for hearing-impaired listeners. *Journal of Rehabilitation Research and Development*, *24*(4), 149–160.

Byrne, D. (1986). Effects of frequency response characteristics on speech discrimination and perceived intelligibility and pleasantness of speech for hearing-impaired listeners. *J. Acoust. Soc. Am.*, *80*(2), 494–504.

Byrne, D. & Dillon, H. (1986). The national acoustic laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear & Hear*, *7*(4), 257–265.

Byrne, D., Parkinson, A., & Newall, P. (1991). Modified hearing aid selection procedures for severe/profound hearing losses. In G. A. Studebaker, F. H. Bess, & L. B. Beck (Eds.), *The Vanderbilt hearing aid report II* (pp. 295–300). Parkton, MD: York Press.

Byrne, D. & Tonisson, W. (1976). Selecting the gain of hearing aids for persons with sensorineural hearing impairment. *Scandinavian Audiology*, *5*, 51–59.

Carney, L. H. (1993). A model for the responses of low-frequency auditory-nerve fibers in cat. *J. Acoust. Soc. Am.*, *93*, 401–417.

Ching, T. Y. C., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *J. Acoust. Soc. Am.*, *103*, 1128–1140.

Dallos, P. & Harris, D. (1978). Properties of auditory nerve responses in absence of outer hair cells. *J. Neurophysiol.*, *41*, 365–383.

Danaher, E. M. & Pickett, J. M. (1975). Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss. *J. Speech Hear. Res.*, *18*, 261–271.

Delgutte, B. & Kiang, N. Y. (1984). Speech coding in the auditory nerve: I. vowel-like sounds. *J. Acoust. Soc. Am.*, *75*, 866–878.

Dillon, H. (1988). Compression in hearing aids. In R. E. Sandlin (Ed.), *Handbook of hearing aid amplification*, volume 1. Boston: College Hill.

Dillon, H. (2001). *Hearing Aids.* New York: Thieme.

Dreschler, W. A. (1988b). Dynamic range reduction by peak clipping or compression and its effects on phoneme perception in hearing-impaired listeners. *Scandinavian Audiology, 17*, 45–51.

Dreschler, W. A., Eberhardt, D., & Melk, P. W. (1984). The use of single-channel compression for the improvement of speech intelligibility. *Scandinavian Audiology, 13*, 231–236.

Evans, E. F. & Wilson, J. P. (1973). The frequency selectivity of the cochlea. In A. R. Moller (Ed.), *Basic Mechanisms in Hearing* (pp. 519–554). New York: Academic.

Franck, B. A., van Kreveld-Bos, C. S., Dreschler, W. A., & Verschuure, H. (1999). Evaluation of spectral enhancement in hearing aids, combined with phonemic compression. *J. Acoust. Soc. Am., 106*, 1452–1464.

Gennaro, S. D., Braida, L., & Durlach, N. (1986). Multichannel syllabic compression for severely impaired listeners. *J. Rehabil. Res. Dev., 23*(1), 17–24.

Glasberg, B. R. & Moore, B. C. J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *J. Acoust. Soc. Am., 79*, 1020–1033.

Glasberg, B. R. & Moore, B. C. J. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scand. Audiol. Suppl., 32*, 1–25.

Glasberg, B. R. & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res., 47*, 103–138.

Johnson, D. H. & Swami, A. (1980). The transmission of signals by auditory-nerve fiber discharge patterns. *J. Acoust. Soc. Am., 68*, 1115–1122.

Kiang, N. Y., Liberman, M. C., & Levine, R. A. (1976). Auditory-nerve activity in cats exposed to ototoxic drugs and high-intensity sounds. *Ann. Otol. Rhinol. Laryngol., 85*, 369–388.

Kiessling, J., Schubert, M., & Archut, A. (1996). Adaptive fitting of hearing instruments by category loudness scaling (scaladapt). *Scand. Audiol., 25*(3), 153–160.

King, A. B. & Martin, M. C. (1984). Is agc beneficial in hearing aids? *British Journal of Audiology, 18*, 31–38.

Knudsen, V. O. & Jones, I. H. (1935). Artificial aids to hearing. *Laryngoscope, 45*, 48–69.

Laurence, R. F., Moore, B. C. J., & Glasberg, B. R. (1983). A comparison of behind-the-ear high fidelity linear hearing aids and two channel compression aids, in the laboratory and in everyday life. *British Journal of Audiology, 17*, 31–48.

Leek, M. R. & Summers, V. (1996). Reduced frequency selectivity and the preservation of spectral contrast in noise. *J. Acoust. Soc. Am., 100*, 1796–1806.

Liberman, M. C. & Dodds, L. W. (1984b). Single-neuron labeling and chronic cochlear pathology. iii. stereocilia damage and alterations of threshold tuning curves. *Hear. Res., 16*, 55–74.

Liberman, M. C. & Mulroy, M. J. (1982). Acute and chronic effects of acoustic trauma: Cochlear pathology and auditory nerve pathophysiology. In R. P. Hamernik, D. Henderson, & R. Salvi (Eds.), *New Perspectives on Noise-Induced Hearing Loss* (pp. 105–135). New York: Raven.

Lindholm, J., Dorman, M., Taylor, B., & Hannley, M. (1988). Stimulus factors influencing the identification of voiced stop consonants by normal-hearing and hearing-impaired adults. *J. Acoust. Soc. Am., 83*(4), 1608–1614.

Lippmann, R., Braida, L., & Durlach, N. (1981). Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. *J. Acoust. Soc. Am., 69*(2), 524–534.

Lybarger, S. F. (1944). U.s. patent application sn 543,278.

Lybarger, S. F. (1978). Selective amplification: A review and evaluation. *J. Am. Audiol. Soc., 3*, 258–266.

McCandless, G. A. & Lyregaard, P. E. (1983). Prescription of gain/output (pogo) for hearing aids. volume 34 (pp. 16–21).

McGrath, M. & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *J. Acoust. Soc. Am., 77*(2), 678–685.

Miller, R. L., Calhoun, B. M., & Young, E. D. (1999). Contrast enhancement improves the representation of /e/-like vowels in the hearing-impaired auditory nerve. *J. Acoust. Soc. Am., 106*, 2693–2708.

Miller, R. L., Schilling, J. R., Franck, K. R., & Young, E. D. (1997). Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibres. *J. Acoust. Soc. Am., 101*, 3602–3616.

Moore, B. C., Laurence, R. F., & Wright, D. (1985). Improvements in speech intelligibility in quiet and in noise produced by two-channel compression hearing aids. *British Journal of Audiology, 19*, 175–187.

Mustafa, K. & Bruce, I. C. (In Press). Robust formant tracking for continuous speech with speaker variability. *IEEE Transactions on Speech and Audio Processing.*

Nedzelnitsky, V. (1980). Sound pressuresin the basal turns of the cat cochlea. *J. Acoust. Soc. Am., 68*, 1676–1689.

Palmer, A. R. (1990). The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *J. Acoust. Soc. Am., 88*, 1412–1426.

Palmer, A. R. & Moorjani, P. A. (1993). Responses to speech signals in the normal and pathological peripheral auditory system. *Prog. Brian Res., 97*, 107–115.

Palmer, A. R., Winter, I. M., & Darwin, C. J. (1986). The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primary like cochlear nucleus neurons. *J. Acoust. Soc. Am., 79*, 100–113.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of vowels. *J. Acoust. Soc. Am., 24*(2), 175–184.

Pickett, J. M. (1980). *The Sounds of Speech Communication.* Austin, TX: Pro-Ed, Inc.

Pickles, A. (1988). *An Introduction to Auditory Physiology* (2nd ed.). New York, NY: Academic Press.

Plomp, R. (1994). Noise, amplification, and compression: considerations of three main issues in hearing-aid design. *Ear & Hearing, 15*(1), 2–12.

Preves, D. A. (1991). Output limiting and speech enhancement. In G. A. Studebaker, F. H. Bess, & L. B. Beck (Eds.), *The Vanderbilt hearing aid report II* (pp. 35–51). Parkton, MD: York Press.

Quatieri, T. F. (2002). *Discrete-Time Speech Signal Processing. Principles and Practice.* Upper Saddle River, NJ: Prentice Hall.

Robertson, D. (1982). Effects of acoustic trauma on stereocilia structure and spiral ganglion cell tuning properties in the guinea pig cochlea. *Hear. Res., 7*, 55–74.

Robertson, D. & Johnstone, B. M. (1979). Aberrant tonotopic organization in the inner ear damaged by kanamycin. *J. Acoust. Soc. Am.*, *66*, 466–469.

Robles, L. & Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiol. Rev.*, *81*, 1305–1352.

Ruggero, M. A. & Rich, N. C. (1991). Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane. *J. Neurosci.*, *11*, 1057–1067.

Sachs, M. B., Bruce, I. C., Miller, R. L., & Young, E. D. (2002). Biological basis of hearing-aid design. *Annals of Biomedical Engineering*, *30*, 157–168.

Salvi, R., Perry, J., Hamernik, R. P., & Henderson, D. (1982). Relationships between cochlear pathologies and auditory nerve and behavioral responses following acoustic trauma. In R. P. Hamernik, D. Henderson, & R. Salvi (Eds.), *New Perspectives on Noise-Induced Hearing Loss* (pp. 165–188). New York: Raven.

Schilling, J. R., Miller, R. L., Sachs, M. B., & Young, E. D. (1998). Frequency shaped amplification changes the neural representation of speech with noise-induced hearing loss. *Hear. Res.*, *117*, 57–70.

Schmiedt, R. A., Mills, J. H., & Adams, J. C. (1990). Tuning and suppression in auditory nerve fibers of aged gerbils raised in quiet or noise. *Hear. Res.*, *45*, 221–236.

Schmiedt, R. A., Zwislocki, J. J., & Hamernik, R. P. (1980). Effects of hair cell lesions on responses of cochlear nerve fibers. i. lesions, tuning curves, two-tone inhibition, and responses to trapezoidal-wave patterns. *J. Neurophysiol.*, *43*, 1367–1389.

Schuknecht, H. F. (1960). Neuroanatomical correlates of auditory sensitivity and pitch discrimination in the cat. In G. L. Rasmussen & W. F. Windle (Eds.), *Neural Mechanisms of the Auditory and Vestibular Systems* (pp. 76–90). Springfield: Thomas.

Schwartz, D., Lyregaard, P., & Lundh, P. (1988). Hearing aid selection for severe-to-profound hearing loss. *The Hear. J.*, *41*(2), 13–17.

Seewald, R., Ramji, K., Sinclair, S., Moodie, K., & Jamieson, D. (1993). *Computer-assited implementation of the desired sensation level method for electroaoustic selection and fitting in children: Version 3.1. Users Manual.* London, Ontario: The University of Western Ontario.

Shaw, E. A. G. (1974). The external ear. In W. D. Keidel & W. D. Neff (Eds.), *Handbook of Sensory Physiology*, volume 5 (pp. 455–490). Berlin: Springer.

Sidwell, A. & Summerfield, Q. (1985). The effect of enhanced spectral contrast on the internal representation of vowel-shaped noise. *J. Acoust. Soc. Am.*, *78*, 495–506.

Simpson, A. M., Moore, B. C. J., & Glasberg, B. R. (1990). Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. *Acta Otalaryngol. Suppl.*, *469*, 101–107.

Sinex, D. G. & Geisler, C. D. (1983). Responses of auditory-nerve fibers to consonant-vowel syllables. *J. Acoust. Soc. Am.*, *73*, 602–615.

Smoorenburg, G. F. (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *J. Acoust. Soc. Am.*, *91*, 421–437.

Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., & Mott, J. (1985). Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. *J. Acoust. Soc. Am.*, *77*, 620–627.

Stone, M. A. & Moore, B. C. J. (1992). Spectral feature enhancement for people with sensorineural hearing impairment: Effects on speech intelligibility and quality. *J. Rehabil. Res. Dev.*, *29*, 39–56.

Stone, M. A. & Moore, B. C. J. (1999). Tolerable hearing-aid delays. i. estimation of limits imposed by the auditory path alone using simulated hearing losses. *Ear Hear.*, *20*, 182–192.

Stone, M. A. & Moore, B. C. J. (2002). Tolerable hearing-aid delays. II. estimation of limits imposed during speech production. *Ear Hear.*, *23*, 325–338.

Stone, M. A. & Moore, B. C. J. (2003). Tolerable hearing-aid delays. III. effects on speech production and perception of across-frequency variation in delay. *Ear Hear.*, *24*, 175–183.

Summerfield, A. Q., Foster, J., Tyler, R. S., & Bailey, P. J. (1985). Influences of formant narrowing and auditory frequency selectivity on identification of place of articulation in stop consonants. *Speech Commun.*, *4*, 213–229.

Summerfield, A. Q. & Haggard, M. P. (1975). Vocal tract normalization as demonstrated by reaction times. In G. Fant & M. Tatham (Eds.), *Auditory Analysis and Perception of Speech* (pp. 115–141). New York: Academic Press.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philos. Trans. R. Soc. Lond. Biol.*, *335*(1273), 71–78.

Summers, V. & Leek, M. R. (1997). Intraspeech spread of masking in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, *101*, 2866–2876.

Tasell, D. J. V., Fabry, D. A., & Thibodeau, L. M. (1987). Vowel identification and vowel masking patterns of hearing-impaired subjects. *J. Acoust. Soc. Am.*, *81*, 1586–1597.

Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. C., & Patterson, R. D. (1984). Auditory filter asymmetry in the hearing-impaired. *J. Acoust. Soc. Am.*, *76*, 1363–1368.

Ulehlova, L., Voldrich, L., & Janisch, R. (1987). Correlative study of sensory cell density and cochlear length in humans. *Hearing Res.*, *28*, 149–151.

Valente, M. & Vliet, D. V. (1990). The independent hearing aid fitting forum (ihaff) protocol. *Trends in Amplification*, *2*(1), 6–35.

Verschuure, J., Maas, A. J., Stikvoort, E., de Jong, R. M., & Dreschler, W. A. (1996). Compression and its effect on the speech signal. *Ear & Hear*, *17*(2), 162–175.

Villchur, E. (1973). Signal processing to improve speech intelligibility in perceptive deafness. *J. Acoust. Soc. Am.*, *53*, 1646–1657.

Walker, G. & Dillon, H. (1982). Compression in hearing aids: An analysis, a review and some recommendations. In NAL *Report No. 90.* Australian Government Publishing Service.

Wiener, F. M. & Ross, D. A. (1946). The pressure distribution in the auditory canal in a progressive sound field. *J. Acoust. Soc. Am.*, *18*, 401–408.

Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B., & Young, E. D. (1998). Effects of high sound levels on responses to the vowel /ϵ/ in cat auditory nerve. *Hear. Res.*, *123*, 61–77.

Young, E. D. & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J. Acoust. Soc. Am.*, *66*, 1381–1403.

Yund, E. & Buckles, K. (1995). Multichannel compression hearing aids: effects of number of channels on speech discrimination in noise. *J. Acoust. Soc. Am.*, *97*(2), 1206–1223.

Zhang, X., Heinz, M. G., Bruce, I. C., & Carney, L. H. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. nonlinear tuning with compression and suppression. *J. Acoust. Soc. Am.*, *109*, 648–670.

Appendix A

MATLAB Code

A.1 Main Function of the MICEFS Algorithm

```
function sigout = micefs(sig, Fs, data_len, zeropad, F1, F2, F3,
F4);

%   MICEFS uses an overlap-add FFT structure to apply multiband
%   compression and spectral enhancement to the input signal.
%
%   This compresses audio using an FFT-based filterbank. The default
%   is a 15-band compressor with filter center-frequencies spaced
%   1/3-octave apart starting at 250 Hz, filter widths ~4 ERBs, fast-
%   acting (near-instantaneous attack, ~60 ms release time constants),
%   with 2:1 compression in all bands.
%
%   A spectral enhancement is applied to the input signal by using a
%   time-varying FIR filter with arbitrary amplitude response evaluated
%   by using the audiogram of a hearing impaired individual and the
%   formant frequencies of the input signal.
```

```
%
%   Much of the specifics of the overlap-add structure were chosen to
%   minimize time-aliasing associated with circular convolution that
%   results from this implementation. If you are going to alter the
%   structure, you can use the distortion out of the compressor with
%   a sinusoid as an input as a measure of merit.
%
%   The system is calibrated so that a unity-amplitude pure tone
%   corresponds to a 100 dB SPL signal. Calibration is important
%   because the gain that gets applied depends on the absolute level
%   out of the filters.
%
%   INPUTS
%   sig: column vector input signal to be compressed.
%   Fs: sampling rate of the input signal; default is 16000.
%   data_len: number of signal data points to include in each FFT
%       frame (not counting zeropadding); default is 128.
%   zeropad: number of zeros appended to the data samples in orders to
%       avoid time aliasing from frequency domain linear filtering.
%   F1, F2, F3 and F4: first four formants of the input signal.
%
%   OUTPUT
%   sigout: column vector of compressed input signal; length is longer
%       than sig because of the extent of the equivalent FIR response
%       of FFT system.
%
%   NON-STANDARD FUNCTION CALLS
%   create_filter, gain_calc, smooth_power
%
```

```
%   Author: Brent Edwards
%   Date: August 2002
%   Modified: Shahab U. Ansari
%   Date: July 2005
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%************************** Set Defaults ******************************
if nargin<3
    data_len = 128;
    frame_len = 512;
end

if nargin<2
    Fs = 16000;
    sig = 10^(-55/20)*sin(2*pi*1000*[1:3*Fs]/Fs)';
    sig(16001:32000) = sig(16001:32000)*10^(35/20);
    plot(sig);
end

if size(sig,2)>size(sig,1)
    sig = sig';
end

% how much the window is advanced for each frame; functions below
% don't work properly if advance is not half of data_len
advance = data_len/2; %64
frame_len = data_len+zeropad;  %512
frame_num = ceil(length(sig)/advance)+1;
```

```
%****************** Filterbank specification ********************
%specify parameters for the filterbank
filter_num = 15; fc_low  = 250; slope = -30; fc_sep = 1/3;

% Displaying the variables
param = sprintf('\nFs = %i\n data_len = %i\n frame_len = %i\n...
    advance = %i\n fc_low = %i\n slope = %i\n fc_sep = 1/3\n'...
    ,Fs,data_len,frame_len,advance,fc_low,slope);
disp(param); disp('Press any key to continue.'); pause

% generate matrix of amplitude response of filters and vector of
% filter center frequencies; filter is a matrix where each row is
% a filter and each column is an FFT bin
[filter fc] = create_filter(filter_num, fc_low, slope, fc_sep, Fs,
frame_len);



%****************************** Windowing ************************
% define various parameters
window = hann(data_len,'periodic');

% Initialize matrix for input signal to be divided into overlapping
% blocks and windowed
win_sig = zeros([frame_len frame_num]);

% each column is a frame;
```

```
% place half of zeropadding before and half after the signal instead
% of all after the signal, which reduces time aliasing from circular
% convolution during gain application. Note: this is only setup
% properly for the condition of advance=data_len/2.
win_sig(zeropad/2+1:zeropad/2+data_len,1) = ...
    [zeros([advance 1]); sig(1:advance)].*window;
    % first frame only half a window data
% Testing delay in impulse response
%     [zeros([20 1]); sig(1:data_len-20)].*window;

for i = 1:frame_num-3
    win_sig(zeropad/2+1:zeropad/2+data_len,i+1) = ...
        sig((i-1)*advance+1:(i-1)*advance + data_len).*window;
end

win_sig(zeropad/2+1:zeropad/2+data_len,frame_num-1) = ...
    [sig((frame_num-2-1)*advance+1:length(sig));...
    zeros([(frame_num-2-1)*advance+data_len-length(sig) 1])].*window;
win_sig(zeropad/2+1:zeropad/2+data_len,frame_num) = ...
    [sig((frame_num-1-1)*advance+1:length(sig));...
    zeros([(frame_num-1-1)*advance+data_len-length(sig) 1])].*window;



%************************ Speech Formants ***************************
%initialize matrix for windowed F1, F2, F3 and F4 values.
win_F1 = zeros([data_len frame_num]); win_F2 = zeros([data_len
frame_num]); win_F3 = zeros([data_len frame_num]); win_F4 =
zeros([data_len frame_num]);
```

```
% Windowed matrix containing formant values is created. Incorporates
% 50% overlap in order align the formant values with input signal
index = 1; for i = 1:frame_num
   win_F1(:,i) = F1(index:(index+data_len)-1);
   win_F2(:,i) = F2(index:(index+data_len)-1);
   win_F3(:,i) = F3(index:(index+data_len)-1);
   win_F4(:,i) = F4(index:(index+data_len)-1);

   if (index+advance) > (length(F1)-data_len)
       index = ((length(F1) - data_len+1));
   else
       index = index + advance;
   end
end

% Hardcoded formants for vowel plus noise
win_F1(:,:) = 500; win_F2(:,:) = 1.7e3; win_F3(:,:) = 2.5e3;
win_F4(:,:) = 3.3e3;



%******************** Frequency Domain Filtering ********************
% create matrix of the FFT of each block of the windowed input signal
% blocks; each column is a frame spectrum
% win_fsig = fft(win_sig,frame_len);
win_fsig = fft(win_sig);

% create power matrix; each column is a frame, each row is the power
```

```
% out of each filter
pow=zeros(size(win_fsig));

pow=(filter.*filter)*(abs(win_fsig).*abs(win_fsig));

%scale power so a sinusoid of full scale gives a power of 1
scale = sum(window.^2)/2;
% pow=(pow/512)/scale;
pow=(pow/frame_len)/scale;



%************************ Compression Setup ********************
% this gives a release time of ~60 ms when measured with a 1kHz tone
% and 2:1 compression everywhere remember to subtract 16ms when
% calculating release time if referencing input timing
smooth = 0.68; pow_est = smooth_power(pow, smooth);

% define gains in dB for 50 dB SPL and 80 dB SPL input, where entry
% G50(k) corresponds to filter(k,:) with center frequency fc(k)
% G50 = [(0-(10*(1-1/2)))*ones([3 1]); (0-(10*(1-1/1)))*ones([3 1]);...
%(0-(10*(1-1/1)))*ones([4 1]);(0-(10*(1-1/2)))*ones([filter_num-10 1])];
% G80 = [(0-(40*(1-1/2)))*ones([3 1]); (0-(40*(1-1/1)))*ones([3 1]);...
%(0-(40*(1-1/1)))*ones([4 1]);(0-(40*(1-1/2)))*ones([filter_num-10 1])];

% 2:1 compression across all filterbands
% G50 = [(0-(10*(1-1/2)))*ones([filter_num 1])];
% G80 = [(0-(40*(1-1/2)))*ones([filter_num 1])];
```

```
comp_ratio = 2/1; %the amount of compression
kneepoint = 40; %input level at which the compression kicks off
Gknee = 0; %gain at kneepoint
G50 = [(Gknee-((50-kneepoint)*(1-1/comp_ratio)))*ones([filter_num
1])]; G80 =
[(Gknee-((80-kneepoint)*(1-1/comp_ratio)))*ones([filter_num 1])];

% 1:1 compression across all filterbands
% G50 = [(0-(10*(1-1/1)))*ones([filter_num 1])];
% G80 = [(0-(40*(1-1/1)))*ones([filter_num 1])];

% 4:1 compression for first 6 filters and no compression for last 6
% filters in filterband
%G50=[(0-(10*(1-1/4)))*ones([6 1]);(0-(10*(1-1/1)))*ones([filter_num-6 1])];
%G80=[(0-(40*(1-1/4)))*ones([6 1]);(0-(10*(1-1/1)))*ones([filter_num-6 1])];

% alternating 2:1 compression across the 15 filterbands
% G50 = [(0-(10*(1-1/2))); (0-(10*(1-1/1))); (0-(10*(1-1/2)));...
%        (0-(10*(1-1/1))); (0-(10*(1-1/2))); (0-(10*(1-1/1)));...
%        (0-(10*(1-1/2))); (0-(10*(1-1/1))); (0-(10*(1-1/2)));...
%        (0-(10*(1-1/1))); (0-(10*(1-1/2))); (0-(10*(1-1/1)));...
%        (0-(10*(1-1/2))); (0-(10*(1-1/1))); (0-(10*(1-1/2)))];
% G80 = [(0-(40*(1-1/2))); (0-(40*(1-1/1))); (0-(40*(1-1/2)));...
%        (0-(40*(1-1/1))); (0-(40*(1-1/2))); (0-(40*(1-1/1)));...
%        (0-(40*(1-1/2))); (0-(40*(1-1/1))); (0-(40*(1-1/2)));...
%        (0-(40*(1-1/1))); (0-(40*(1-1/2))); (0-(40*(1-1/1)));...
%        (0-(40*(1-1/2))); (0-(40*(1-1/1))); (0-(40*(1-1/2)))];
```

```
%****************** Gain Calc for Comp. and Enhanc. ****************
% calculate the gain from each smoothed power estimate and the gain
% parameters; as usual, each column is a frame and each row is an FFT bin
win_gain = gain_calc(pow_est, G50, G80, frame_len, Fs, fc, win_F1,
win_F2, win_F3, win_F4, kneepoint);

% multiply the gain matrix by the spectrum matrix, then inverse FFt to
% get block time-domain output
win_sigout = real(ifft(win_fsig.*win_gain));



%************************** Over-and-Add ***************************
%initialize output vector
sigout=zeros([frame_num*advance+frame_len 1]);

% finally, generate compressed signal vector with overlap-add
for i = 1:frame_num
    sigout((i-1)*advance + 1:(i-1)*advance + frame_len) = ...
    sigout((i-1)*advance + 1:(i-1)*advance + frame_len)+win_sigout(:,i);
end
```

A.2 Function for Creating Filterbank

```
function [filter, fc] = create_filter(filter_num, fc_low, slope,
fc_oct_sep, Fs, frame_len)
```

```
%   CREATE_FILTER creates a filterbank for the multiband compressor. Each
%   filter has a frequency region where the response is unity. The filters
%   roll off above and below with a slope specified by the user. The extent
%   of the rolloff is an octave above and below the unity-gain region.
%
%   The bandwidth of the unity-gain sections are set to four times the
%   equivalent rectangular bandwidth (erb) of that center frequency. This
%   allows for reasonable overlap between the filters when the center
%   frequencies have 1/3-octave separation, minimizing spectral ripple for
%   tonal stimuli. The lowest band is a lowpass filter with cutoff frequency
%   set to the first "center" frequency. The spacing of the filters is
%   determined by the specified octave separation between center frequencies.
%   The amount of overlap between bands depends on this and the bandwidth of
%   the filters.
%
%   INPUTS
%   filter_num: number of filters in the filterbank.
%   fc_low: the lowest filter center frequency (actually the kneepoint for
%       the lowpass).
%   slope: rolloff of the filters outside of the unity gain region, specified
%       in dB/octave (must be negative).
%   fc_oct_sep: frequency separation between filter center frequencies,
%       specified in octaves.
%   Fs: sampling rate of the system.
%   frame_len: length of the frame for the input to the FFT (Nyquist is at
%       frame_len/2+1).
%
%   OUTPUTS
```

```
%   filter: a matrix specifying the amplitude response of the filter_num
%       filters. Each row is a filter and each column is a frequency
%       corresponding to the bin number of the FFT signal, i.e., #1 is DC,
%       #2 is Fs/frame_len, #3 is 2*Fs/frame_len, etc. the rightmost columns
%       are a mirror image of the leftmost columns (excluding DC).
%   fc: a vector of filter center frequencies.
%
%   Author: Brent Edwards
%   Date: August 2002
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%calculate the frequency separation between FFT bins
bin_width = Fs/frame_len;

% calculate the center frequency of each filter
for i=1:filter_num
    fc(i) = fc_low*2^((i-1)*fc_oct_sep);
end

% make sure maximum center frequency is not greater than the Nyquist
% frequency
if (max(fc)>Fs/2)
    fprintf('\n ERROR IN FILTERBANK GENERATION \n');
    pause
end

% calculate the bandwidth of the unity gain of each filter
erb = 24.7*(4.37*fc/1000+1);
```

```
% initialize filter matrix
filter = zeros([filter_num frame_len]);

% specify amplitude response of first filter, a special case since
% it's a lowpass
indx = 1:ceil(fc(1)/bin_width)+1; filter(1,indx) =
ones(size(indx));

indx_above = max(indx)+1:2*max(indx); filter(1,indx_above) =
10.^(log2((indx_above-1)*bin_width/fc(1))*slope/20);

% specify the amplitude response for the rest of the filters, storing
% each response in a new row in the matrix
for i =2:filter_num
    indx = [floor((fc(i)-2*erb(i))/bin_width)+1:min(ceil((fc(i)+...
    2*erb(i))/bin_width)+1,frame_len/2+1)];
    filter(i,indx) = ones(size(indx));
    indx_above = max(indx)+1:min(2*max(indx),frame_len/2+1);
    filter(i,indx_above) = 10.^(log2((indx_above-1)*bin_width/...
    ((max(indx)-1)*bin_width))*slope/20);
    indx_below = max(ceil(min(indx)/2),2):min(indx)-1;
    filter(i,indx_below) = 10.^(-log2((indx_below-1)*bin_width/...
    ((min(indx)-1)*bin_width))*slope/20);
end

% add the mirror image of the filter responses for application on
% the FFT-domain
filter(:,frame_len/2+2:frame_len) = filter(:,frame_len/2:-1:2);
```

A.3 Function for Smoothing Signal Power for Compression

```
function pow_est = smooth_power(pow, release_coef);

%   SMOOTH_POWER smooths the power estimate out of filters with instantaneous
%   attack and specified release times. This implements a power estimation
%   process that has a near-instantaneous attack time and a release time
%   dependent on release-coef. The smoothed power estimate for each filter
%   is equal to the instantaneous power of the frame if that power is greater
%   than the smoothed power estimate. Otherwise, the smoothed power estimate
%   for that frame is updated using a single-pole IIR filter.
%
%   Time constants are estimated using a sinusoid input that starts at 55 dB SP
%   increases to 90 dB SPL, then returns to 55 dB SPL. The attack time is the
%   time the output takes to get within 3 dB of steady-state after the increase
%   from 55 to 90 dB SPL; the release time is the time the output takes to get
%   within 4 dB of steady-state after the decrease from 90 to 55 dB SPL. Don't
%   forget to ignore the group delay of the system if you measure this.
%
%   INPUTS
%   pow: a matrix of power calculations from each filter for each frame. Each
%       column is a frame, each row is a filter.
%   release_coef: IIR coefficient used for the power smoothing on release. The
%       larger the number, the longer the release time.
%
%   OUTPUTS
%   pow_est: a version of pow with the power for each filter smoothed over time
```

```
%       with the same dimensions as pow.
%
%   Author: Brent Edwards
%   Date: August 2002
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

frame_num  = size(pow,2);   % determine number of frames to process

pow_est = pow;  % not necessary, but makes what's going on easier to understand

% smooth the power estimate on a frame-by-frame basis. The power estimate for
% the first frame is simply equal to the instantaneous power for that frame.
for i=2:frame_num
    %select filters where current power is less than the smoothed estimate
    k = find(pow(:,i)<pow_est(:,i-1));
    pow_est(k,i) = release_coef*pow_est(k,i-1)+(1-release_coef)*pow(k,i);
end
```

A.4 Function for Calculating Gain

```
function win_gain = gain_calc(pow_est, G50, G80, frame_len, Fs,
    fc, win_F1, win_F2, win_F3, win_F4, kneepoint, gain_type)

%   GAIN_CALC calculates the gain for each filter band then determines the
%   gain response for the FFT signal. The gain for each filter's center
%   frequency is calculated for each frame according to specified I/O
```

```
%   functions for each filter. Then, the gain is interpolated/extrapolated
%   to each of the frequency bins in the FFT signal.
%
%   The gain function for each filter is defined by the gains for 50 and 80
%   dB SPL input levels. The gain function is linear on a dB-dB scale,
%   calculated by linearly interpolating (in the dB domain) based on the G50
%   and G80 values. The compression ratio is calculated as,
%   CR = 30/(30 - [G50 - G80]).
%
%   kneepoint defines the input level below which gain remains constant, i.e.,
%   compression stops and the filter response is linear. It is set here to 40
%   dB SPL. The compression gain is also modified to ensure that gain is not
%   lower than 0 dB. So, at high levels the filter response is linear pass-
%   through. An alternate implementation of this is to set a highlevel
%   kneepoint specifying the input level above which the gain does not change,
%   but this is not implemented here.
%
%   INPUTS
%   pow_est: matrix of power used to specify the gain. Each column is power
%       from a different frame and each row is power from a different filter.
%   G50: a vector of gains for 50 dB SPL input to each filter.
%   G80: a vector of gains for 80 dB SPL input to each filter.
%   frame_len: size of the FFT frame.
%   Fs: sampling rate of system.
%   fc: vector of filterbank center frequencies.
%   win_F1,win_F2,win_F3,win_F4: windowed first four formants.
%   kneepoint: input level at which the compression kicks off.
%   gain_type: defines the amplification scheme
%
```

```
%   win_gain: matrix of amplitude gain to be applied by the compressor, where
%       each column is a different frame and each row is a different FFT bin;
%       the gain in each row is symmetric excluding the first (DC) bin.
%
%   Author: Brent Edwards
%   Date: August 2002
%   Modified: Shahab U. Ansari
%   Date: July 2005
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%***************************** Initialization ******************************
bin_width = Fs/frame_len; filter_num = size(pow_est,1); frame_num
= size(pow_est,2);
NF = Fs/2; %Nyquist frequency

% Scale the power so that a power intensity of 1 corresponds to 100 dB SPL
% (recall in the main compressor function that a sinusoid with unity amplitude
% is defined as having a level of 100 dB SPL, and the power estimate was scaled
% such that a unity-amplitude sinusoid produces an intensity of 1).
pow_dB = 100+10*log10(pow_est);


%************************** Compression ********************************
% Calculate gain using matrices to simplify the interpolation/extrapolation
% computations. An equivalent scalar expression is given as,
% gain_dB = G50 + (pow_dB-50)*(G80-G50)/(80-50);
matrix1 = (G50-50*(G80-G50)/(80-50))*ones([1 frame_num]); matrix2
= (G80-G50)/(80-50)*ones([1 frame_num]); gain_dB =
```

```
matrix1+matrix2.*pow_dB; clear matrix1 matrix2

%compression kneepoint in dB, linear below this point; make sure gain is >=0
% kneepoint = 40;
Gknee = (G50 + (kneepoint-50)*(G80-G50)/(80-50))*ones([1
frame_num]); i = find(gain_dB>Gknee); gain_dB(i) = Gknee(i);

% interpolate gain in filters to FFT bins
indx_low = ceil(fc(1)/bin_width); indx_high =
floor(fc(filter_num)/bin_width);

win_gaindB = zeros([frame_len/2 frame_num]);
win_gaindB(indx_low:indx_high,:) = interp1q(fc',gain_dB,
[indx_low*bin_width:bin_width:indx_high*bin_width]');

%extrapolate gain in filters to FFT bins
if indx_low > 1
    win_gaindB(1:indx_low-1,:) = ones([indx_low-1, 1])*gain_dB(1,:);
end

if indx_high < frame_len/2
    win_gaindB(indx_high+1:frame_len/2,:) =
        ones([frame_len/2-indx_high, 1])*gain_dB(filter_num,:);
end


%******************** Spectral Enhancement ************************
%Hearing loss profile
load('modelaudiogram');
```

```
if (strcmp(gain_type, 'HG')) % Half gain amplification
    f = [0:(Fs/2)/(frame_len/2):Fs/2-1]';

    fs = 100:200:8e3;
    H3dB = interp1(BFkHz,THSH,fs/1e3,'linear','extrap')/2;
    H = 10.^(interp1(fs,H3dB,f,'linear','extrap')/20);

    H3dB = interp1(BFkHz,THSH,f/1e3,'linear','extrap')/2;
    H = 10.^(H3dB/20);
elseif (strcmp(gain_type, 'NALR')) % NAL_R gain calculation
    f = [0:(Fs/2)/(frame_len/2):Fs/2-1]';

    H3FA=sum(interp1(BFkHz,THSH,[500,1000,2000]/1e3,'linear'))/3;
    X = 0.15*H3FA;

    % Prescribed gain
    freq = [250,500,1000,2000,3000,4000,6000];
    gdB = [-17,-8,1,-1,-2,-2,-2];

    % Thresholds from hearing loss profile
    thr = interp1(BFkHz,THSH,freq/1e3,'linear','extrap');

    % NAL-R gain
    H_dB = X + 0.31.*thr + gdB;
    H = 10.^(interp1(freq,H_dB,f,'linear','extrap')/20);
else
    H = 0;
```

```
end;

% For filter subplots with formant noise
% fig = figure;

for i = 1:frame_num
    F1 = round(mean(win_F1(:,i)));
    F2 = round(mean(win_F2(:,i)));
    F3 = round(mean(win_F3(:,i)));
    F4 = round(mean(win_F4(:,i)));

    % Formants with noise
%     std = 100;
%     std = 200;
%     std = 300;
%     F1 = round(mean(win_F1(:,i)))+round(randn(1,1)*std);
%     F2 = round(mean(win_F2(:,i)))+round(randn(1,1)*std);
%     F3 = round(mean(win_F3(:,i)))+round(randn(1,1)*std);
%     F4 = round(mean(win_F4(:,i)))+round(randn(1,1)*std);

    % Half-gain rule (CEFS)
    gain_F2 = interp1(BFkHz,THSH,F2/1e3)/2;

    if (strcmp(gain_type, 'MICEFS'))
        % test11
        gain_F3 = gain_F2+5;
        gain_F2 = gain_F2-10;
    end;
```

```
if (strcmp(gain_type, 'CEFS')) % FIR filter for CEFS
    b = fir2(512, ...
        [0 (F2-50)/NF F2/NF F3/NF 1], ...
        [1 1 10^(gain_F2/20) 10^(gain_F2/20) sqrt(10^((gain_F2)/20))])';

    [H,f] = freqz(b,1,frame_len/2+1,Fs);
end;

if (strcmp(gain_type, 'MICEFS')) % FIR filter for MICEFS
    if ((2*F1+14*F2)/16 > (F1+15*F2)/16 ||...
            (F1+15*F2)/16 > (14*F2+2*F3)/16 ||...
            (14*F2+2*F3)/16 > (13*F2+3*F3)/16 ||...
            (13*F2+3*F3)/16 > (2*F2+14*F3)/16 ||...
            (2*F2+14*F3)/16 > (F2+15*F3)/16 ||...
            (F2+15*F3)/16 > (15*F3+F4)/16 ||...
            (15*F3+F4)/16 > (14*F3+2*F4)/16)
        continue;
    end

    % test11
    %*******
    b = fir2(512, ...
        [0 (2*F1+14*F2)/16/NF (F1+15*F2)/16/NF (14*F2+2*F3)/16/NF...
            (13*F2+3*F3)/16/NF (2*F2+14*F3)/16/NF (F2+15*F3)/16/NF...
            (15*F3+F4)/16/NF (14*F3+2*F4)/16/NF   1], ...
        [1 1 10^(gain_F2/20) 10^(gain_F2/20) 10^((gain_F2-10)/20)...
            10^((gain_F2-10)/20) 10^(gain_F3/20)  10^(gain_F3/20)...
            10^((gain_F3-10)/20) sqrt(10^((gain_F3-10)/20))])';
```

```
        [H,f] = freqz(b,1,frame_len/2+1,Fs);
    end;

    if (strcmp(gain_type, 'MICEFS'))
        % Combined gain for compression and spectral enhancement
        win_gaindB(:,i) = win_gaindB(:,i) + 20*log10(abs(H(2:end)));
    elseif (strcmp(gain_type, 'CEFS')) % Gain CEFS without compression
        win_gaindB(:,i) = 20*log10(abs(H(2:end)));
    else %Gain for half-gain or NAL-R without compression
        win_gaindB(:,i) = 20*log10(abs(H));
    end;

    % for frequency response plots with noisy formants
%     plot(f,db(abs(H)));
%     hold on;
%     wysiwyg
%     paperfig2(fig,12)

end

% figure;
plot(f,db(abs(H))); hold on;
% plot([1700 1700],[-5 35],'k:');
% plot([2500 2500],[-5 35],'k:');
xlabel('Frequency (Hz)'); ylabel('Amplitude (dB)');
title('Spectral enhancement filter');

% Replicate the frequency response for negative frequencies
win_gain = 10.^(win_gaindB/20); win_gain = [zeros([1 frame_num]);
```

```
win_gain; win_gain(frame_len/2-1:-1:1,:)];
```